Human macular gene expression

by

Daniel Mark Hornan

submitted for the degree of

Doctor of philosophy in Genetics

Division of Molecular Genetics

Institute of Ophthalmology, University College London

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The human macula is essential for precise vision. It contains many more cone photoreceptors than the peripheral retina, especially in the fovea. Cones are known to express specific opsins and other proteins that form part of the phototransduction cascade.

However, relatively little is known about retinal macular gene expression compared with the rod-rich peripheral retina. I obtained human donor eyes and used foveo-macular and macular punches and sections of peripheral retina to study differential gene expression. I combined multiple microarray experiments with quantitative PCR, statistical, and bioinformatic analyses.

I identified several known and previously unidentified retinal genes that are more abundant in the macula. I went on to characterize proteins encoded by histone deacetylase 9 and the *morpheus* gene family. Both were expressed in the human macula, especially in the photoreceptors. Several other genes also provided insight into the mechanisms of precise vision and its maintenance. Genes identified by this approach are excellent candidates for macular disease.

Contents

List of Figures

List of Tables

List of abbreviations

aa	amino acids
ABI	Applied Biosystems International
ACTB	beta actin
ad	autosomal dominant
AEBSF	4-(2-Aminoethyl)benzenesulphonyl fluoride
AMD	age-related macular degeneration
apoE	apolipoprotein
ar	autosomal recessive
AREDS	age related eye disease study
ARM	age related maculopathy
ARP	acidic ribosomal phosphoprotein
b	base
BBS	Bardet-Biedl syndrome
BLAST	Basic Local Alignment Search Tool
BMP	bitmap
bp	base pairs
BSA	bovine serum albumin

CCD	charge coupled device
cDNA	complimentary DNA
cGMP	cyclic guanosine monophosphate
CNV	choroidal neovascularization
Crx	cone rod homeobox
Cy3	cyanine 3
Da	Dalton
DAPI	4',6-Diamidino-2-phenylindole
DART-PCR	Data Analysis for Real-Time PCR
dbEST	EST database at NCBI
dCTP	deoxyribose cytosine triphosphate
DIC	differential interference contrast
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribose nucleic acid
dpi	dots per inch
dUTP	deoxyuracil triphosphate
ECACC	European Collection of Cell Cultures
ECL	enhanced chemiluminescence
EGF	epidermal growth factor
EIF	eukaryotic translation initiation factor 2 alpha kinase 4
ELISA	enzyme-linked immunosorbent assay
EMBOSS	European Molecular Biology Open Software Suite
EOG	electro-oculogram

ERG	electro-retinogram
EST	expressed sequence tag
EtBr	ethidium bromide
FBS	fetal bovine serum
FCH	FER/CIP4 homology
FISH	fluorescence *in situ* hybridization
GA	geographic atrophy
GABA	gama-amino butyric acid
GAP	GTPase-activating protein
GCL	ganglion cell layer
GEO	Gene Expression Omnibus
GFAP	glial fibrillary acid protein
GFP	green fluorescent protein
GST	glutathione S-transferase
GTP	guanosine triphosphate
H	histone
HDAC	histone deacetylase
HPLC	high performance liquid chromatography
HRP	horesradish peroxidase
IACUC	Institutional Animal Care and Use Committee
ILM	inner limiting membrane
INL	inner nuclear layer
IPL	inner plexiform layer

IPTG	Isopropyl beta-D-thiogalactopyranoside
IRBP	interphotoreceptor binding protein
JPEG	Joint Photographic Experts Group
KLH	keyhole limpet hemacyanin
LB	Luria-Bertani
LCR	low-copy repeat
LGN	lateral geniculate nucleus
MBP	maltose binding protein
MBS	m-Maleimidobenzoyl-N-hydroxysuccinimide ester
MEGAP	mental disorder-associated GAP protein
MeOH	methanol
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
NDS	normal donkey serum
NFL	nerve fibre layer
NPIP	nuclear pore complex interacting protein
OMIM	Online Mendelian Inheritance in Man
ONL	outer nuclear layer
OPL	outer plexiform layer
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDE	phosphodiesterase
PDF	portable document forma

PMT	photomultiplier tube
PNA	peanut agglutinin
qPCR	quantitative PCR
RDS	retinal degeneration, slow
RHOGAP	Rho GTPase activating protein
RNA	ribonucleic acid
ROM	rod outer segment membrane protein
RP	retinitis pigmentosa
RPE	retinal pigment epithelium
RPM	revolutions per minute
rRNA	ribosomal RNA
RS	retinoschisin
RT-PCR	reverse transcriptase PCR
SAG	s-antigen
SAGE	serial analysis of gene expression
SDS	sodium dodecyl sulphate
SFD	Sorsby fundus dystrophy
SMA	statistics for microarray analysis
SQL	structured query language
SRGAP	SLIT-ROBO Rho GTPase activating protein
SRNVM	sub-retinal neovascular membrane
SSC	sodium chloride - sodium citrate
Taq	Thermus aquaticus

TB	transfer buffer
TGF	transforming growth factor
TIFF	tagged image file format
TIMP	tissue inhibitor of metalloproteinase
tRNA	transfer RNA
UBC	ubiquitin C
VEGF	vascular endothelial growth factor
VMD	vitelliform macular dystrophy
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactoside
XLRS	X-linked retinoschisis

Chapter 1

Introduction

The human eye (Figure 1.1) is a highly specialized organ of vision. It subserves the process by which light energy from the environment produces changes in specialized nerve cells in the retina – the rod and cone photoreceptors. These changes result in nerve impulses that are subsequently relayed via the optic nerve to the brain where the information is consciously appreciated as vision. All the other structures in the eye are secondary to this process. Some structures are part of the system necessary for focusing and transmitting the light onto the retina – for example the cornea, iris, ciliary body and lens. Other structures are necessary for nourishing and supporting the tissues of the eye – for example the lacrimal (tear) apparatus, choroid and aqueous outflow system.

The human eye is approximately spherical, with a diameter of 2.5cm and a volume of 6.5ml. More accurately, it is part of two spheres: a smaller one anteriorly, the cornea, that has a greater curvature than the sclera, which constitutes the larger sphere. The cornea constitutes one sixth of the circumference of the eye and has a diameter 15.5mm. The remaining five-sixths is formed by the sclera, which has a diameter of 23mm.

The human eye is situated in the anterior part of the orbit, closer to the lateral than the medial wall and nearer the roof than the floor. The eye is made up of three basic layers or coats, often known as tunics. These are the fibrous (corneo-scleral) coat, the uvea or uveal tract (composed of choroid, ciliary body, and iris), and the neural layer (retina). The coats surround the contents – the lens and the transparent media (aqueous humour

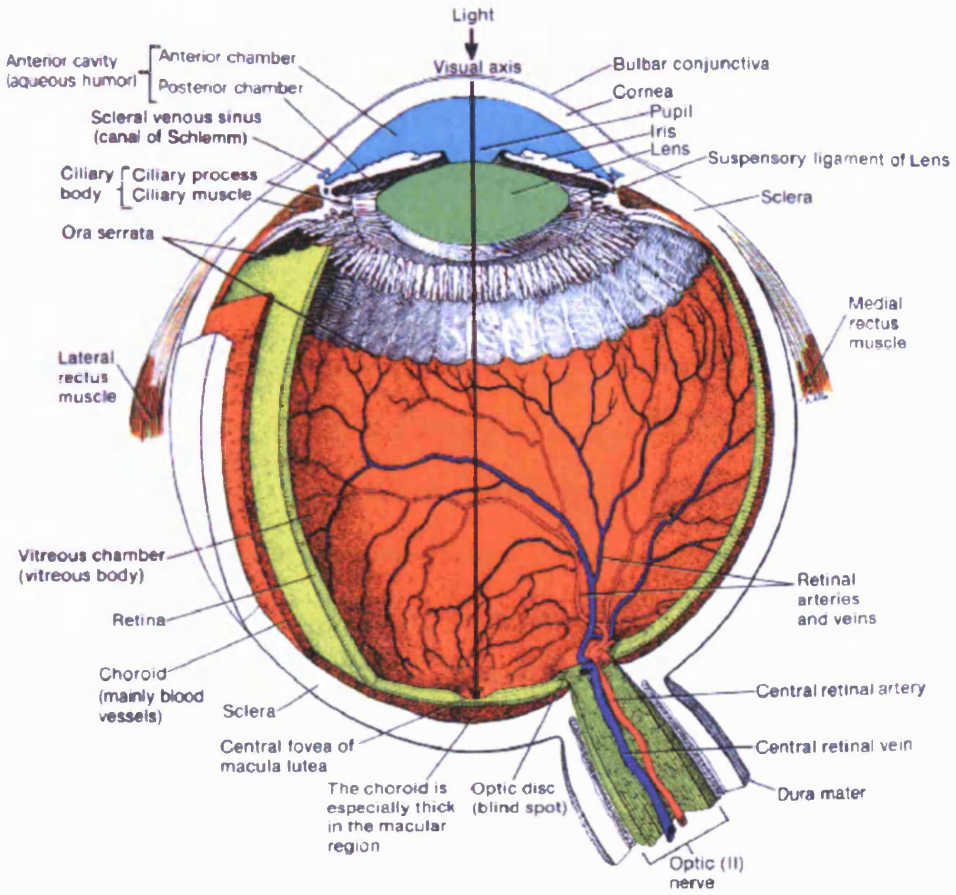


Figure 1.1: Transverse section of the left eye (superior view). Adapted from `academia.hixie.ch`

and vitreous jelly). The cornea and sclera together form a tough envelope that protects the ocular tissues. This fibrous coat also provides important structural support for the intra-ocular contents and for attachment of extra-ocular muscles.

1.1 Retina

The human retina is the innermost layer of the eye. It is the plane in which the image falls after it is focused by the eye's optical system. The retina is responsible for converting information from the image of the external environment into neural impulses that travel to the brain. It is the light-sensitive intima of the eye, from which the optic nerve emanates.

The vertebrate retina consists of two main layers: an inner neurosensory retina and an outer epithelium – the retinal pigment epithelium (RPE). These two layers can be traced embryologically to the inner and outer layers of the optic cup. In the adult they are continuous anteriorly with the epithelial layers over the ciliary processes and the posterior surface of the iris. Between the neural retina and the RPE is a potential space, the sub-retinal space, across which the two layers adhere. The neural retina is only firmly attached at its anterior termination, the ora serrata, and at the optic nerve head border. Adhesion between the neural retina and RPE is normally maintained by electrostatic forces, negative pressure and proteoglycans.

The retina is bound by Bruch's membrane externally and the vitreous internally. It is continuous with the optic nerve posteriorly, the site of exit of ganglion cell axons from the eye.

The central retina is a 5 to 6mm diameter circular zone situated between the superior and inferior temporal retinal arteries. It is distinct from the peripheral retina in the predomination of cone rather than rod photoreceptors. Also, the inner retina (INL, IPL, GCL, NFL) is thicker due to the increased density of cone-synapsing second order neurones and is characterized by the presence of more than a single layer of ganglion cells. In the pit of the fovea, a small depression at the centre of the retina, the densely packed hexagonal mosaic of cone photoreceptors displaces all other layers laterally.

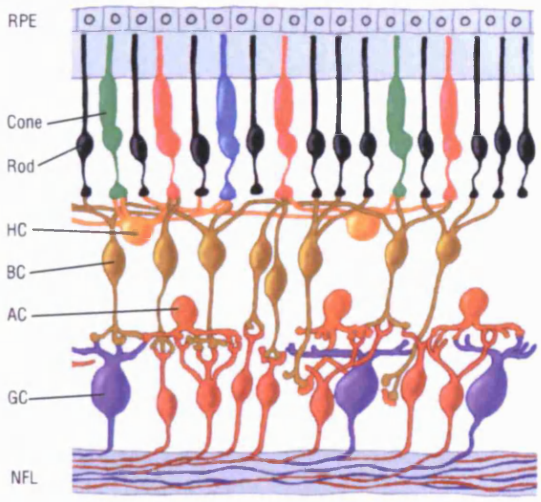


Figure 1.2: Schematic diagram of the vertebrate retina. *NFL* = nerve fibre layer. *GCL* = ganglion cell layer. *AC* = amacrine cell. *BC* = bipolar cell. *HC* = horizontal cell. *RPE* = retinal pigment epithelium. Adapted from `www.americanscientist.org`.

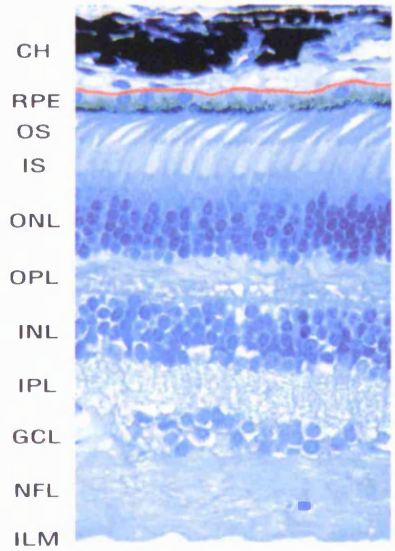


Figure 1.3: Light micrograph showing a vertical section of human retina. *ILM* = inner limiting membrane. *NFL* = nerve fibre layer. *GCL* = ganglion cell layer. *IPL* = inner plexiform layer. *INL* = inner nuclear layer. *OPL* = outer plexiform layer. *ONL* = outer nuclear layer. *IS* = photoreceptor inner segment. *OS* = photoreceptor outer segment. *RPE* = retinal pigment epithelium. *CH* = choroid. Bruch's membrane is highlighted in red. Adapted from `medocs.ucdavis.edu`.

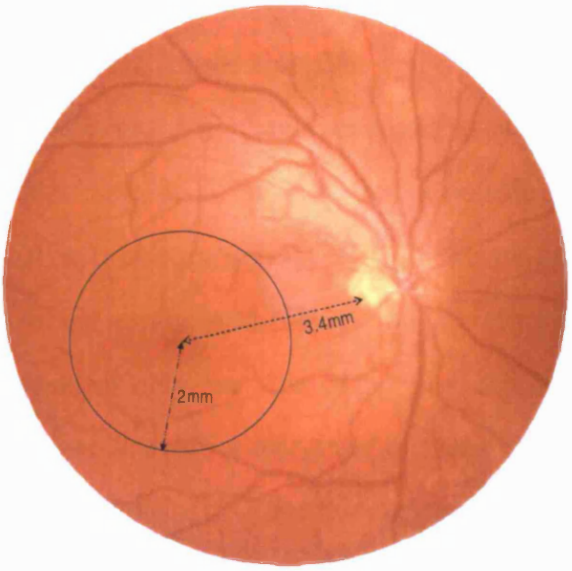


Figure 1.4: Fundus photograph of a normal human retina. The macula (circled) appears as a darker area of pigmentation 11.8° temporal to the optic disc.

The human *macula* is a yellow spot (*macula lutea*) surrounding the fovea, defined as 3mm of intense pigmentation, surrounded by a 1mm wide zone of less pigmentation [Polyak, 1941]. The macula is located 3.4mm temporal to the optic disc margin. One degree of visual angle is equal to 288μm on the retina, without correction for shrinkage [Drasdo and Fowler, 1974]. The macula is yellow partly as a result of yellow screening xanthophyll carotenoid pigments (xeaxanthin and lutein) in the cone axons. This may serve to act as a short wavelength filter protecting against UV irradiation. Through an ophthalmoscope, the macular pigmentation appears as a darker area compared with the peripheral retina (see Figure 1.4).

The fovea is defined as the central 1.5mm of the macula and the foveola is the central 0.35-mm wide zone, consisting of a depression surrounded by slightly thickened margins. The foveola is where cone photoreceptors are concentrated at maximum density to the exclusion of rods. The inner retinal layers in the margins of the pit (clivus) are displaced laterally and the outer retina consists purely of cones. The foveal retina is avascular and relies on the choriocapillaris for nutritional support.

The optic disk lies 3mm medial to the centre of the macula. There are no normal retinal layers in this zone, hence the blind spot. At the optic disk, the ganglion cell axons

from the retina pierce the sclera to enter the optic nerve. This pale pink/whitish area is 1.8mm in diameter with a slightly raised rim. The central retinal vessels emerge at the centre of the optic disk, pass over the rim, and radiate out to supply the retina.

The peripheral retina is the remainder of the retina outside the central retina. The distance from the optic disk to the ora serrata is 23-24mm on the temporal aspect and approximately 18.5mm on the nasal aspect. The peripheral retina is 110-140μm in thickness, rich in rods, and possesses only one layer of ganglion cell bodies.

The ora serrata is the dentate anterior margin of the sensory retina. At this transition zone, the neuroretina is continuous with the columnar non-pigmented epithelial cells in the pars plana of the ciliary body. The ora serrata is around 1mm closer to the corneoscleral limbus on the nasal than on the temporal side.

The retina is divided, for descriptive purposes, into nasal and temporal halves by a vertical line through the fovea. The optic nerve head is often used as central point to describe the retina as having supero- and infero-nasal and supero- and infero-temporal quadrants. The area of the retina is approximately 1250mm^2 and varies in thickness from 100μm in the periphery to 230μm near the optic nerve head.

1.1.1 Retinal pigment epithelium

The Retinal pigment epithelium (RPE) is a continuous monolayer of cuboidal or columnar epithelial cells that extends from the margin of the optic nerve head to the ora serrata, where it is continuous with the pigment epithelium of the ciliary body. This cell layer has many physical, metabolic, biochemical, optical and transport functions, which play a critical role in the normal visual process. These include: maintaining adhesion of the neurosensory retina, providing a selectively permeable barrier between the choroid and neurosensory retina, phagocytosis of rod and (to a lesser extent) cone outer segments, synthesis of interphotoreceptor matrix, absorption of light and reduction of scatter (thus improving image resolution), and transport as well as storage of metabolities and vitamins (especially vitamin A). The complex morphology of this neuroectoderm-derived epithe-

lium reflects these multiple functions.

The RPE cells vary in size and shape depending on age and location, being more columnar in the central retina and more flattened in the peripheral retina. The basal aspect of the cells lies on Bruch's membrane and their apical surface is intimately associated with the photoreceptor outer segments. When examined 'en face' they form a highly organized hexagonal pattern of homogeneously sized cells.

The RPE has low regenerative capacity in the normal eye. Therefore, cell loss is accommodated by hyperplasia of adjacent cells. Therefore, in older eyes the regular hexagonal array is lost and a heterogeneous mixture of sizes and shapes is more evident. The number of RPE cells per eye varies from 4-6 million.

1.1.2 Neural retina

The neural retina is a thin transparent layer of neural tissue, which in life has a red/purple tinge due to the presence of visual pigments. *Post mortem* and in fixed specimens the neural retina is opaque white and often detached from the underlying RPE.

Light stimuli are converted into neural impulses in the neural retina. Some degree of processing then occurs locally before transmission of information to the brain via the ganglion cell axons in the optic nerve. The neural retina consists of several cell types. As well as neurones, other cell types include glial cells, vascular endothelium and microglia. The three principal neurone cell types that relay impulses generated by light are photoreceptors, bipolar cells, and ganglion cells – their activity is modulated by horizontal cells, amacrine cells, and possibly by non-neuronal elements such as the neuroglia. It is the culmination of this neural processing concerning the visual image that is eventually transmitted to the brain along the optic nerve. Retinal cells are arranged in a highly organized manner, and in histological sections appear as distinct layers that include three layers of nerve cell bodies and two layers of synapses (Figure 1.3).

The neural retina acts both as a *transducer* and a *comparator*. Photoreceptors convert light energy from absorbed photons into electrical energy in the form of neural impulses.

The circuitry between the photoreceptors and the retinal ganglion cells serves to compare the level of neural activity in different photoreceptors – this ability to detect contrast is a fundamental part of vision.

1.1.3 Photoreceptors

There are two types of photoreceptor in the human eye: rods and cones. They are situated on the outer aspect of the retina. There are approximately 120 million rods and 6 million cones in the human eye. Rods are responsible for sensing contrast, brightness, and motion, while cones subserve fine spatial resolution and colour vision. The density of rods and cones varies in different regions of the retina. The central 0.5mm of the fovea, which is 1.5mm in total diameter, contains only cone photoreceptors. Away from the centre the spatial density of cones falls rapidly from a maximum of $\approx$150,000/mm^2 to $\approx$7,000/mm^2 at 3mm retinal eccentricity. The spatial density of rods is zero in the centre of the retina, but rises quickly to reach a peak of $\approx$150,000/mm^2 at an eccentricity of 5-7mm, beyond which it steadily declines [Rodieck, 1998].

Each photoreceptor consists of a long narrow cell with an inner and outer segment joined by a connecting stalk consisting of a modified cilium (Figure 1.5). These inner and outer segments are separated from the cell body by an outer limiting membrane formed by Müller cells. The nucleus is situated in the outer nuclear layer of the retina and axons pass into the outer plexiform layer where they form synaptic terminals with bipolar cells and horizontal cells. The outer segments of rods and cones are shaped as their names imply. They contain the visual pigments that are responsible for absorption of light and initiation of *phototransduction* impulse.

Rods

Rods are long (100pm) slender cells whose outer segment contains the visual pigment rhodopsin, which has a maximal spectral sensitivity of 500nm. Rhodopsin is contained within membrane-bound disks. There are up to 1000 disks per cell, each of which is

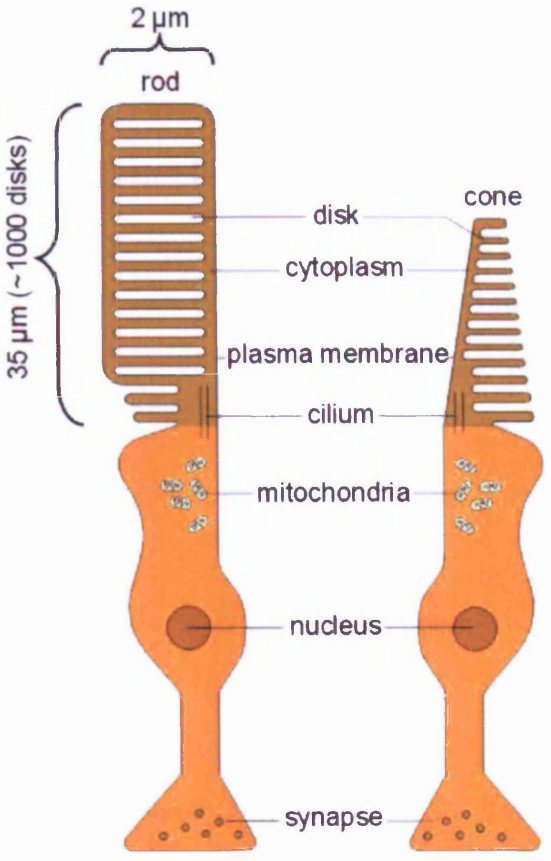


Figure 1.5: Schematic drawing of rod and cone photoreceptor cells. The primary events of phototransduction take place in the outer segment of the cell, shown in brown.

10-15nm thick. The disks are enclosed by a single membrane. Each outer segment is 1-1.5μm in width and 25μm long. The disks are produced at the base of the outer segment and over the course of 10 days travel to the rod tips, which are enclosed by the apical microvilli of the RPE. Here they are phagocytosed by the RPE cells in a circadian manner – rod disks are thought to be predominantly shed in the early morning in humans [Young, 1967], shortly after the onset of light in all species examined [Bernstein et al., 1984].

The rods are separated by a modified extracellular ground substance containing inter-photoreceptor binding protein (IRBP) within a sialic acid and proteoglycan matrix. The inner half of the inner segment is known as the myoid, the outer half the ellipsoid (3μm in length).

The ellipsoid is connected to the outer segment by a modified cilium consisting of nine doublet microtubules without a central pair. The connecting cilium represents the embryological vestige of the ciliated neuroepithelial cells that line the primitive optic ventricle. The cilium acts as a conduit for metabolites and lipids between the inner and outer segments. The remainder of the ellipsoid contains numerous mitochondria, indicative of the high metabolic activity of these cells.

The myoid region contains numerous organelles, including Golgi apparatus, smooth endoplasmic reticulum, microtubules, and glycogen – evidence of a metabolically and synthetically active cell.

Metabolic disturbances in the photoreceptors may lead to disturbances in visual pigment or photoreceptor membrane synthesis and therefore specific abnormalities in vision. In addition dehydration of the interphotoreceptor matrix can lead to retinal detachment due to loss of adhesive forces between the outer segments and the RPE.

Cones

Daylight vision relies exclusively on cones, which generate precise vision due to their density in the foveal region [Hendrickson, 1994]. In most diurnal animals, two spectrally distinct cone types exist – one maximally sensitive to short wavelengths and one to long

wavelengths. However, in Old World primates, apes and humans, a third cone type exists. The three types are generally referred to as red, green, and blue (or long, medium, and short wavelength) cones. Cone outer segments are generally shorter than rods and generally conical (6μm at the base, 1.5μm at the tip) – hence the term cones. However, in the fovea cones are long, slender, and tightly packed. The cone disks are not surrounded by a plasma membrane in the same manner as rods but are in free communication with the interphotoreceptor matrix. Cone disks have a greater lifespan than rods and are thought to be shed shortly after the cessation of light [Bernstein et al., 1984].

Cones are surrounded by the long villous melanin-containing apical processes of the RPE and are about 60-75μm in length. The outer segment is connected to the mitochondria-rich ellipsoid region (containing around 600 mitochondria per cell) of the inner segment by a cilium similar to that described in rods. The cell body of the cone can be identified histologically in the outer nuclear layer because of its large, pale-staining nucleus and perinuclear cytoplasm.

1.1.4 Photoreceptor connections

The cell bodies of rods and cones are connected by an inner fibre to specialized expanded synaptic terminals, known as *spherules* and *pedicles* respectively. These synapse with bipolar and horizontal cells and contain many highly specialized presynaptic vesicles.

Rod *spherules* are further out (towards the sclera) than the cone *pedicles* and are deeply indented by bipolar and horizontal cell processes (*telodendria*). A specialized region known as a synaptic ribbon is present between two adjacent nerve fibres. Horizontal cell *telodendria* penetrate deeply into the rod *spherule*, whereas the bipolar cell dendrites have a shallower penetration. Up to five processes may be embedded in one *spherule*. There is no apparent contact between rod *spherules*; cone *pedicles* may be connected by gap junctions.

Cone *pedicles* are broader than rod *spherules* and have a pyramidal shape. In the cone *pedicle* there are up to 12 indentations, each of which contains three neuronal terminals.

The central process in each triad is a midget bipolar cell dendrite. The lateral processes in the triad are horizontal cell (*telodendria*) that may be involved in several triads on a cone *pedicle*. Therefore, there may be up to 25 synaptic ribbons in each *pedicle*. Each cone is usually contacted by all the horizontal cells (four to six) in the immediate field. Each pedicle also has numerous shallow indentations or synapses with flat diffuse bipolar cells.

1.1.5 Bipolar cells

The retina contains approximately 35 million bipolar cells, which comprise several functional and morphologic subtypes. They are the neurones primarily responsible for transmitting signals from photoreceptors to ganglion cells. Their cell bodies lie in the inner nuclear layer and are oriented radially, parallel to the photoreceptors. Single or multiple bipolar cell dendrites pass outwards to synapse, principally with photoreceptors but also with horizontal cells, while their single axon passes inwards to synapse with ganglion and amacrine cells.

In the foveal region, the ratio of cones:bipolar cells:ganglion cells is 1:1:1, whereas in the peripheral retina one bipolar cell receives stimuli from up to one hundred rods. In intervening regions, the ratio corresponds to the decreasing visual acuity present in the mid-peripheral retina. Summation of stimuli is a crucial factor in the sensitivity of the rod system to low levels of illumination.

Human bipolar cells can be subdivided into nine morphological subtypes: one type of rod bipolar and eight types of cone bipolars. The cone bipolars can in turn be subdivided into five types of diffuse cone bipolars and three types of midget bipolars.

Rod bipolars

Rod bipolar cells have a receptive field or dendritic tree that is small in the central retina (15μm wide, 10-20 rods) and larger in the peripheral retina (30μm, 30-50 rods). These represent 20% of all bipolar cells and are most dense around the fovea. In the periphery, rod bipolars contact up to 50 rods and synapse with AII amacrine cells; only rarely do

they synapse directly with diffuse ganglion cells.

Diffuse cone bipolars

Diffuse cone bipolar cells are concerned with converging information from many cones. Their dendrites fan out up to 100μm to end in clusters of between five and twenty cone pedicles. The overlap of adjacent cells of this type is extensive in the peri-foveal region.

Midget bipolars

Invaginating midget bipolar cells are the smallest of the bipolars. Their dendrites penetrate the base of a single cone pedicle to form the central element in a triads. The near 1:1 ratio of midget bipolar cells:cones decreases peripherally. Midget bipolar dendrites synapse with amacrine cells and midget ganglion cells.

Flat midget bipolar cells are similar to invaginating midget bipolar cells except their dendrites do not invaginate deeply into the cone pedicle. Potentially, most cones can be in contact both types of midget bipolar cells as well as the diffuse type.

Blue cone-specific bipolar cells have been described in primates and humans [Kolb et al., 1992] and appear to make invaginating contact with only the blue cones in their territory.

1.1.6 Ganglion cells

The cell bodies of most ganglion cells are located in the innermost nucleated layer of the retina -- the ganglion cell layer situated between the nerve fibre layer and the inner plexiform layer (Figure 1.2). Ganglion cells are the last link in the retinal component of the visual pathway. Their axons form the nerve fibre layer on the innermost surface of the retina and synapse with cells in the lateral geniculate nucleus (LGN) of the thalamus.

Ganglion cell axons form bundles that are both separated and ensheathed by glial cells. These bundles leave the eye to form the optic nerve. On exiting through the *lamina cribrosa*, the axons become myelinated by oligodendrocytes.

Up to seven layers of ganglion cell bodies are present in the central retina, where the ganglion cell layer is 6O-80μm thick. There is as few as one layer of cells in the peripheral retina (10-20μm). There are approximately a million ganglion cells in the retina so there are about 100 rods and four to six cones per ganglion cell.

While they are functionally diverse, ganglion cells are all characterized morphologically by a large cell body, abundant Nissl substance (arrays of rough endoplasmic reticulum), and large Golgi apparatus. Ganglion cells are classified into different types on the basis of cell body size, dendritic tree spread, branching pattern, and branching level in the inner plexiform layer. Some ganglion cells in the macula may contain yellow (xanthophyll carotenoid) pigment in the cytoplasm.

Ganglion cell impulses are received primarily from bipolar cells and amacrine cells via axo-dendritic and axo-somatic synapses, the former occurring predominantly in the inner plexiform layer where their dendrites repeatedly branch to form the 'dendritic tree'. Thee form and size of the ganglion cell dendritic tree varies considerably and is correlated with location in the retina and therefore function (receptive field size).

Due to their morphological diversity, ganglion cells (up to 25 types in mammalian and 18 types in human retinas) can be classified into categories α , β and γ , or X, Y, and W types, predominantly based on research in the cat. In the cat, functional correlates (contrast, spatial resolution) are better understood than in primates, which have been less well investigated (Rowe, 1991; Kolb et al, 1992).

Midget ganglion cells

Midget ganglion cells synapse with one or more amacrine cells and a single midget bipolar cell and therefore usually one cone. The dendritic tree is about 5-10μm in diameter in the central retina; however, this increases 10-fold at 2mm eccentricity and attains a maximum of over 100μm in the periphery. Neighbouring midget ganglion cell dendritic trees do not overlap but form mosaics. In humans, midget ganglion cells are also known as P-cells because they project to the parvocellular layer of the lateral geniculate nucleus (LGN).

Diffuse (parasol) ganglion cells

Diffuse ganglion cells include a large synaptic field from all types of bipolars except midget bipolar cells. They occur in the central retina and their cell bodies are 8-16μm in diameter with 30-70μm dendritic fields, these being smaller nearer the fovea than in the periphery. Diffuse ganglion cells are also known as M-cells because they project to the magnocellular layer of the LGN. The finding that midget ganglion cells synapse exclusively with the midget bipolar cells (and that both are common near the fovea) provides the anatomical basis for the observation of small receptive fields and high visual acuity in this region.

1.1.7 Horizontal cells

Horizontal cells derive their name from the extensive horizontal extensions of their cell processes. There are two distinct morphological varieties in the retina of most species, of which the cat is the most extensively studied. Type A horizontal cells are large axonless cells with stout dendrites that contact only cones. Type B horizontal cells have smaller bushier dendritic trees that contact cones exclusively, but in addition have axons of up to 300μm in length that end in an extensive arborization that is postsynaptic only to rods. Type A cells have much larger receptive fields than type B. In primates it appears that the two types of horizontal cell, HI (approximates to type B) and HII (approximates to type A), both possess axons. More recently a third type (HIII) has been described in the human retina by Kolb et al. [1994].

Each rod has connections with at least two horizontal cells and each cone with three or four horizontal cells of each type. In primates, the stout dendrites of each HI cell body contact around seven cones near the fovea (the dendritic tree covering 15μm). This number increases to as many as 18 further from the fovea (the dendritic tree covering 80-100μm). The axon from each HI cell passes laterally and terminates up to 1 mm away in a thickened axon terminal in (up to 100) rod spherules.

HII dendritic trees are more spidery and contact about twice as many cones. Their

axons are generally shorter (100-200μm) and contact cone pedicles by small wispy terminals. HII cell bodies are located primarily in the outer part of the inner nuclear layer and have few distinctive cytoplasmic organelles except the crystalloids – a series of densely stacked tubules with associated ribosomes. HII processes ramify in the outer plexiform layer close to the cone pedicles.

The overlap between horizontal cells is considerable and any one area of retina may be served by up to 20 horizontal cells. Horizontal cells have an integrative role in retinal processing and release inhibitory neurotransmitters, mainly gama-amino butyric acid (GABA). Recent evidence suggests that there is some colour-specific wiring for the three types of horizontal cells in the human retina [Ahnelt and Kolb, 1994].

1.1.8 Amacrine cells

Amacrine cells are association neurones that were originally thought to lack axons. However, recent studies have shown that some do indeed possess an axon [Mariani, 1990]. They are located in the inner aspect of the inner nuclear layer and are distinguishable from bipolar cells in this layer as a result of their larger size (12μm) and oval shape. Amacrine cells display a remarkable degree of morphological (and pharmacological) diversity. There are at least 25 different types in the monkey and human retina. Their cell body is usually flask shaped and the numerous dendritic processes of these cells ramify and terminate predominantly in the synaptic complexes formed by bipolar and ganglion cell processes, namely the inner plexiform layer.

The shape of amacrine cell dendritic fields is highly variable. They can be divided into subtypes on several criteria such as the stratification of their dendrites in the inner plexiform layer or their shape for example diffuse, starburst, and stratified. Diffuse types can cover narrow fields (approximately 25μm wide), their fibres being cone shaped. Other types may spread their axon-like processes several millimeters. Amacrine cells may also be classified on the basis of their neurotransmitters. They may be GABAergic and dopaminergic or can release acetylcholine – indicating, together with their morphology,

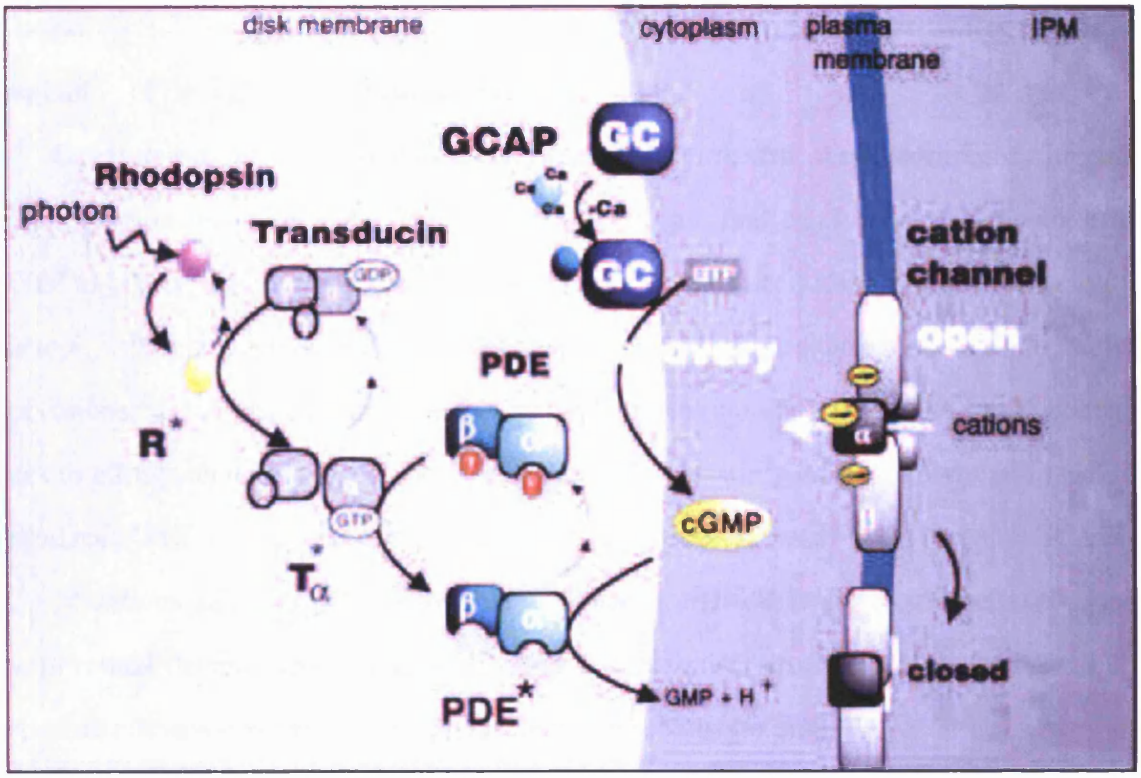


Figure 1.6: Phototransduction cascade.

that these cells play a role in (inhibitory) modulation of signals reaching ganglion cells.

1.2 Phototransduction

The visual pigment of the rod photoreceptors, rhodopsin (R), consists of the transmembrane-spanning protein opsin linked to the chromophore 11-cis-retinal at Lys 296. Rhodopsin is specifically localised to photoreceptor outer segments in the disc (but also plasma) membrane. Photon absorption by the chromophore leads to its isomerization to the all-trans configuration along with accompanying comformational changes. This form of rhodopsin, R*, is catalytically active and binds the G-protein transducin (T). Activated transducin (T*) in turn activates a membrane-associated phosphodiesterase (PDE).

In the photoreceptor outer segment, cation channels are directly gated by cGMP and control the influx of ions across the plasma membrane. The hydrolysis of cGMP by PDE results in a change in cGMP-gated channel conformation resulting in channel closure. This decreases the conductance of the plasma membrane to cations and causes *hyperpo-*

larization and therefore a *decrease* in glutamate neurotransmitter release, which results in signaling of the light stimulus to adjacent neurones.

Calcium ions play a key role in the recovery of the dark state of photoreceptors through the regulation of guanylate cyclase (GC), the enzyme that catalyzes the conversion of GTP to cGMP. Calcium ion concentration is high in the dark ($\approx$500nM) and GC activity is low. After photoactivation, closure of the plasma membrane channels reduces the influx of cations, including calcium. However, the sodium/calcium/potassium exchanger continues to efflux calcium ions and as a result their concentration decreases, activating GC to produce cGMP via the calcium-inactivated guanylate cyclase activating protein (GCAP).

Mutations in many proteins related to the phototransduction cascade are associated with retinal degenerative diseases. These are summarised in Figure 1.7 and those with specific relevance to the macula are discussed in sec:macdisease.

1.2.1 Cones *versus* rods

In cones, the molecules of the biochemical cascade are similar to rods. However, they occur for the most part, on infoldings of the plasma membrane rather than on discrete discs (see Figure 1.5). Cone opsin molecules and other membrane proteins are able to diffuse from infolding to infolding, expanding the number of transducin molecules an activated opsin can encounter and the number of PDE molecules activated transducin can encounter. The most notable difference between cone and rod *response* is that the cone response is biphasic, consisting of a hyperpolarization phase (similar to that in rods) followed by an undershoot (instead of a slow re-polarization) before returning to the baseline. Rods saturate in bright light whereas cones do not.[Rodieck, 1998]

Topological differences between rods and cones represents a major anatomical disparity between the two types of cell, yet we do not yet fully understand its functional significance. One interesting possibility is that the enormous increase in stability of the rhodopsin molecule is aided in some way by the sealing-off of the discs. Perhaps it is advantageous to expose the amino-terminal region of rhodopsin not to the extracellular

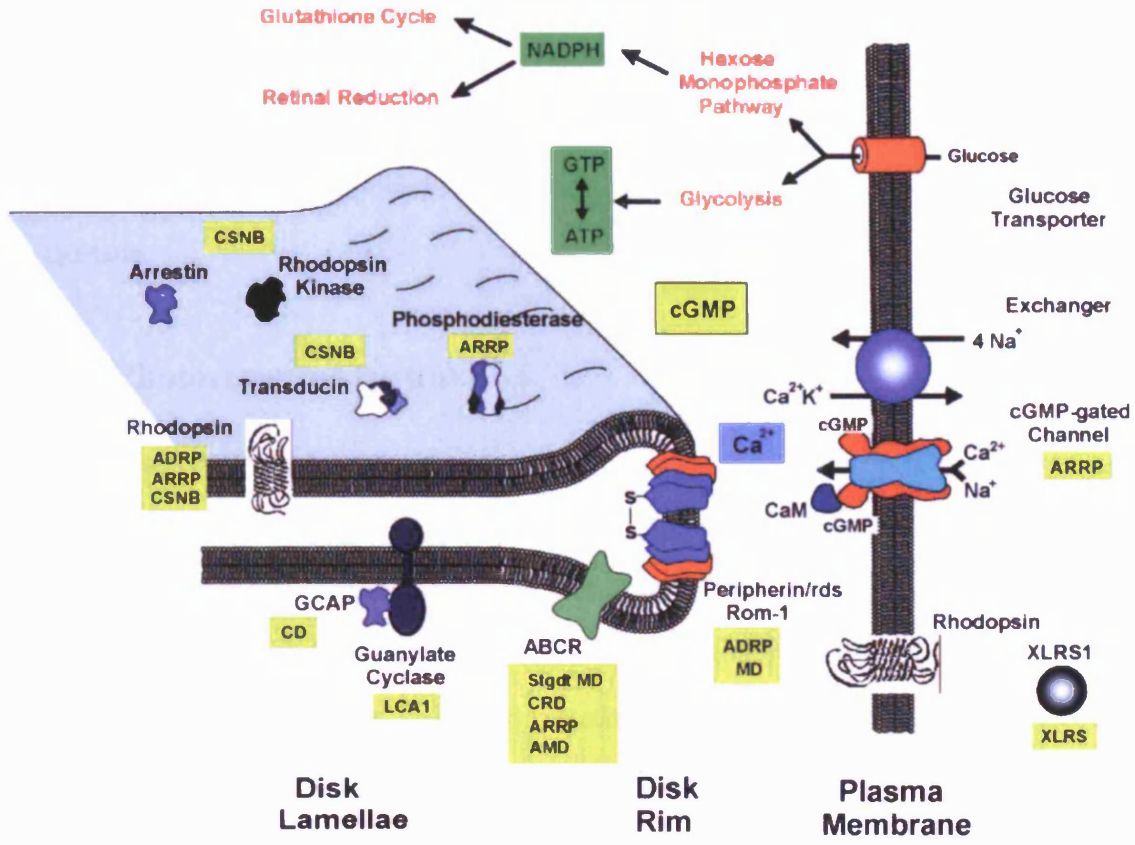


Figure 1.7: Phototoreceptor outer segment proteins linked to retinal disease. ADRP = autosomal dominant retinitis pigmentosa. ARRP = autosomal recessive retinitis pigmentosa. XLRP = x-linked retinitis pigmentosa. CSNB = congenital stationary night blindness. CD = cone dystrophy. CRD = cone-rod dystrophy. LCA = Leber's congenital amaurosis. MD = pattern macular dystrophy. Stgdt MD = Stargardt macular dystrophy. XLRS = X-linked retinoschisis.

solution, but instead to the intra-discal medium. For cones, the greatly increased surface-to-volume ratio that is possible with patent sacs may be important, as this speeds the rate of change of intracellular ion concentrations mediated by plasma membrane fluxes. This is likely to be especially important for calcium ions because of the negative feedback to the cascade.

1.2.2 Photoreceptor turnover

The lipids and proteins that comprise the discs and sacs of photoreceptor outer segments eventually deteriorate. The photoreceptor needs to perform continual maintenance during its life span. Rods periodically shed old discs from the distal tip of the outer segment, and continually generate new foldings at the base of the outer segment, that in due course pinch off to form discs. Thus, rod outer segments are not of uniform age, but rather the newer components 'move up' as they achieve seniority. A similar process of renewal and shedding occurs in cones. However, because the entire membrane is continuous, protein components of different ages are distributed along the length of the cone outer segment.

The existence of quantitative as well as qualitative differences between rods and cones implies that there must be differences in gene expression between them. Furthermore, because of both the increased cone concentration in the macula and the fact that the macula is specialized for precise vision, it follows that there are differences in gene expression between the macula and the peripheral retina. These differences in gene expression can provide molecular clues to understanding both the physiology and pathology of the macula.

1.3 Macular gene expression

Relatively little is known about gene expression in the human macula. This is partly due to the difficulty in obtaining fresh human tissue. A number of expressed sequence tags (ESTs) have been identified from enriched and non-enriched central retinal cDNA

libraries (ncbi.nlm.nih.gov). These ESTs have contributed to the number of genes known to be expressed in the macula, although many of these are likely not to be differentially expressed compared with the peripheral retina. Northern analysis has been used to compare gene expression between central and peripheral retina. For example, ten randomly sequenced clones from a non-normalized foveal cDNA library were used to probe RNA from central and peripheral retina. This confirmed that there are significant differences in gene expression [Bernstein et al., 1996]. More recently, techniques such as serial analysis of gene expression (SAGE) and using subtracted libraries have been used to study macular gene expression.

1.3.1 Serial analysis of gene expression

Serial analysis of gene expression (SAGE) is a relatively new technique that determines the frequency of expression of a large number of genes in a given tissue or cells [Velculescu et al., 1995]. The method is based on two principles.

Firstly, a short (9-20 base) nucleotide sequence that is sufficiently specific to discriminate different transcripts from each other. For example, a 9-base pair sequence can distinguish approximately 250,000 transcripts. Given that the estimated number of human genes is lower than 100,000, unique 9-base pair sequences can theoretically provide expression information on all human genes.

Secondly, concatenation of short sequence tags allows efficient analysis of transcripts by sequencing them within a clone, followed by sequence comparison with the nucleotide database to identify the expressed sequences. The number of times that a sequence of a given gene is found within the isolated clones represents the frequency of the transcript.

In SAGE, double stranded cDNA is synthesized from mRNA of the tissue or cells by using a biotinylated oligo(dT) primer. The cDNAs are cleaved with a restriction endonuclease that recognizes a specific sequence of four or more bases in the cDNA molecule. The most 3' part of the cleaved biotinylated cDNA is bound to strepavidin beads. This assures that the site binding to the strepavidin is closest to the poly(A) tail. The cleaved

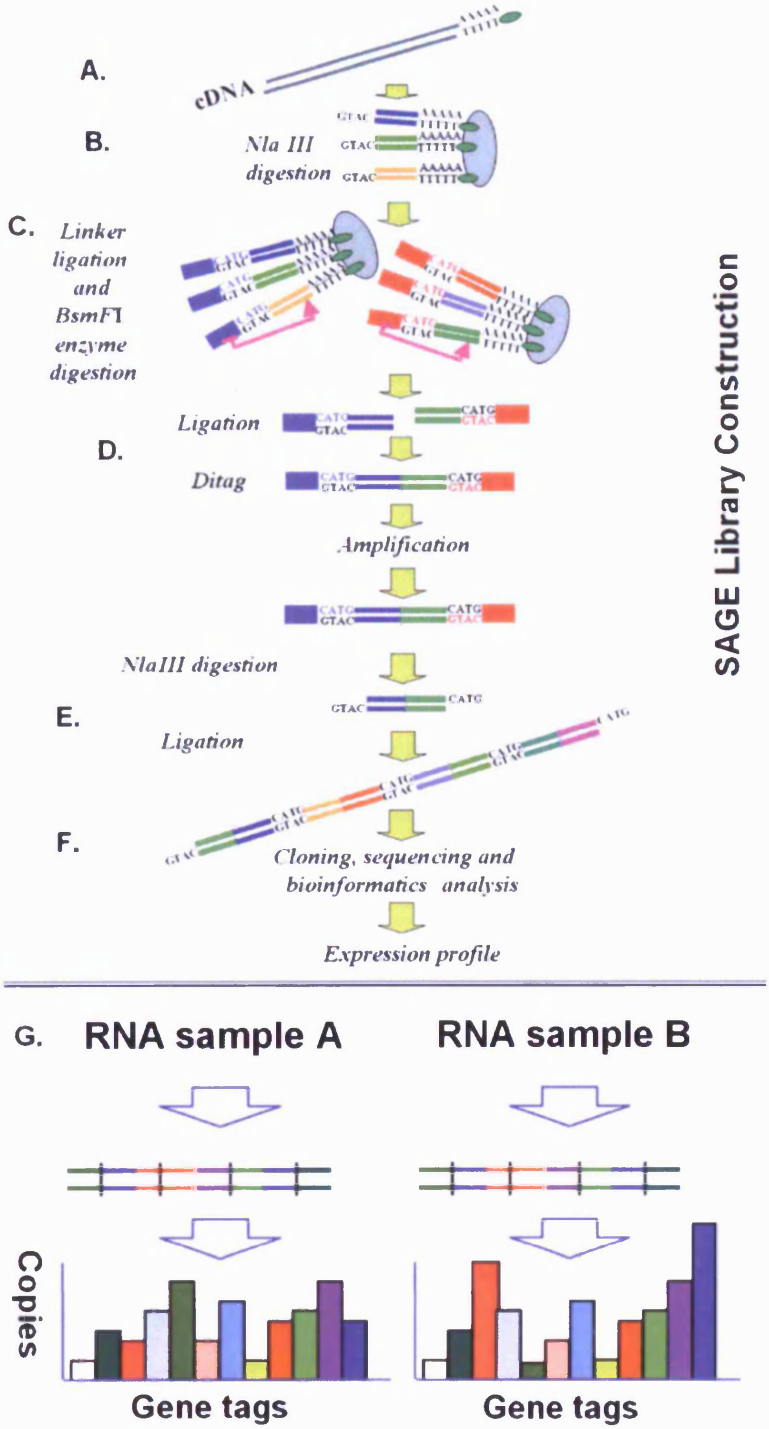


Figure 1.8: Serial analysis of gene expression (SAGE) library construction. (A) An RNA population is reverse transcribed to cDNAs using oligo-T primers attached to magnetic beads. (B) cDNAs are collected and digested with the restriction endonuclease *NlaIII*. (C) Linkers containing sequence recognized by *BsmFI* are ligated to the digested cDNAs. Sequence tags are released from the beads by *BsmFI* digestion (*BsmFI* cuts at a fixed distance downstream from its recognition site). (D) Released DNA tags are ligated together to form ditags. (E) Ditags are amplified and then digested with *Nla III* to remove the linkers. (F) Ditags are ligated together to form a concatemer which is then clones into a plasmid vector to generate a SAGE library. The identity and abundance of tags is deduced from DNA sequence analysis of plasmid clones of concatemated ditags. (G) Relative abundance of gene expression – between genes within the same RNA population or between samples – is deduced by counting sequence tags.

cDNAs are split into two aliquots and ligated, via the 'sticky end' left by enzymatic cleavage, to one of two linkers that contain a type IIS restriction site (tagging enzyme).

Type IIS restriction endonucleases cleave at a defined distance, up to 20bp away from their asymmetric recognition sites. The linkers are designed so that cleavage of the ligation products with the tagging enzyme leads to the release of the linker along with a short fragment of cDNA. The two pools of released tags are ligated to each other by the created blunt ends. This is followed by PCR, which serves to amplify the tag sequences and provides orientation and punctuation of the tag sequences. The amplified product contains the two tags (ditags), which are linked tail to tail and are flanked by sites for the anchoring enzyme. This results in 4bp punctuation per ditag. Then, the ditags are released from the PCR products by cleavage with an anchoring enzyme. Finally, released ditags are concatenated by ligation, cloned and sequenced.

Recently, SAGE was used to compare the peri-macular region (6mm) with peripheral retina [Sharon et al., 2002]. This provided additional evidence for significant quantitative differences in gene expression between central and peripheral retina. Sharon et al. [2002] constructed SAGE libraries using two eyes and sequenced the gene tags. 24 tags were reported to be significantly over-represented in 'macula' (HMAC2-SAGE) compared with peripheral retina (HPR2-SAGE). However, 36 gene tags were also reported to be significantly differentially expressed between the two peripheral retinas (HPR1/2-SAGE) used in the study.

1.3.2 Macular disease

As well as eliciting important mechanisms in precise vision, studying differential expression in the macula may be important to understanding degenerative disease. There are a number of Mendelian retinal degenerations (cone dystrophies, cone-rod dystrophies and macular dystrophies) for which no causative mutation has yet been found (`www.sph.uth.tmc.edu/Retnet`). Monogenic dystrophies are excellent starting places for determining gene function in the retina. Molecular genetic differences between

the peripheral retina and the macula are also likely to be of importance to the much more common multifactorial disease age-related macular degeneration (AMD).

The inherited macular dystrophies are a heterogeneous group of disorders that are characterised by loss of central vision and atrophy of the macula and underlying RPE. Monogenic macular dystrophies are a useful source of insight into the molecular pathology of the macula.

Stargardt disease, ABCA4

Stargardt macular dystrophy is the most common inherited macular dystrophy with a prevalence of 1 in 10,000 and an autosomal recessive mode of inheritance. It shows a very variable phenotype with respect to age of onset and severity. Most cases present with central visual loss in the early teens and there is typically macular atrophy with white flecks at the level of the RPE (see Figure 1.9. Fluorescein angiography classically reveals a dark or masked choroid. This is believed to be secondary to excess lipofuscin accumulation in the RPE [Fish et al., 1981]. It has been demonstrated histologically that the number of photoreceptor cells is reduced in the presence of increased quantities of lipofuscin in the RPE, leading to the proposal that autofluorescent material may accumulate prior to cell death [Dorey et al., 1989].

The causative gene of Stargardt disease has been identified as ABCA4 [Allikmets et al., 1997]. Subsequently, mutations in ABCA4 have also been implicated in other disorders, including retinitis pigmentosa (RP)[Martínez-Mir et al., 1998] and cone-rod dystrophy (CORD) [Maugeri et al., 2000]. ABCA4 encodes a transmembrane rim protein located in the discs of rod and foveal cone outer segments, that is involved in ATP-dependent transport of retinoids from photoreceptors to the RPE [Weng et al., 1999; Molday et al., 2000]. Failure of this transport results in deposition of the major lipofuscin fluorophore, N-retinylidene-N-retinylethanolamine, in the RPE [Weng et al., 1999]. It is proposed that this accumulation may be deleterious to the RPE, with consequent secondary photoreceptor degeneration.

Figure 1.9: Stargardt disease. Fundus photograph showing white flecks at the level of the RPE at the posterior pole. There is early macular atrophy.

Best disease, VMD2

Best disease is a dominantly inherited macular dystrophy that is characterised by a round or oval yellow sub-retinal macular deposit (see Figure 1.10). The yellow material is gradually resorbed over time, leaving an area of RPE atrophy and often sub-retinal fibrosis. The electro-retinogram (ERG) is normal but the electro-oculogram (EOG) shows a very reduced or absent light rise, indicating that there is widespread dysfunction of the RPE [Deutman, 1969]. In common with Stargardt disease, histopathology shows accumulation of lipofuscin throughout the RPE [Weingeist et al., 1982]. Although the ophthalmoscopic abnormality is usually confined to the macular region, this evidence of more widespread retinal involvement is in common with the majority of inherited macular dystrophies described to date.

Best disease shows very variable expressivity and the visual prognosis is surprisingly good, with most patients retaining reading vision into the fifth decade of life or beyond. Most individuals carrying mutations in the vitelliform macular dystrophy 2 (VMD2) gene on chromosome 11q13 [Petrukhin et al., 1998] have an abnormal EOG, but the macular appearance may be normal in some [Mohler and Fine, 1981]. The protein product of

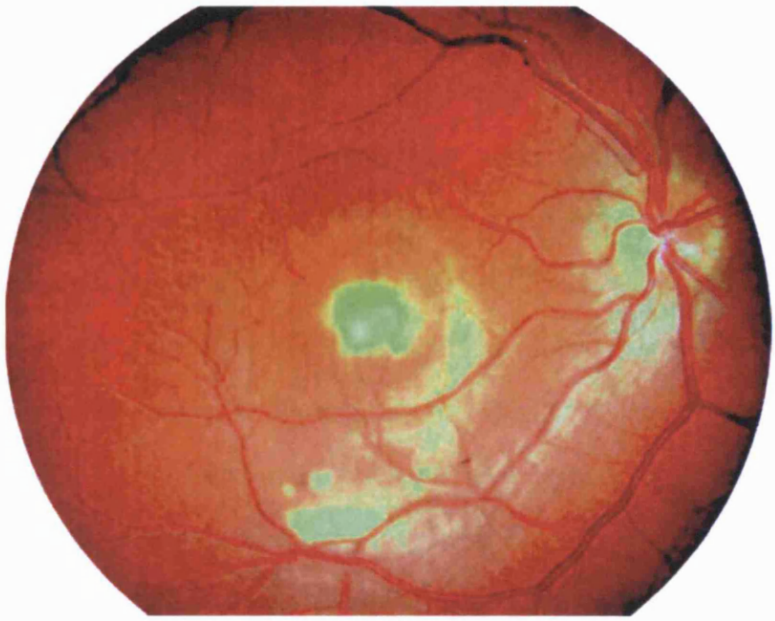


Figure 1.10: Best disease. Fundus photograph of showing a partially resorbed yellow subretinal macular deposit.

VMD2, bestrophin, has been localised to the basolateral plasma membrane of the RPE, where it forms a component of a chloride channel [Marmorstein et al., 2000]. This regulates fluid transport across the RPE, and it has been suggested (following optical coherence tomography) that impaired fluid transport in the RPE secondary to abnormal chloride conductance may lead to accumulation of fluid and/or debris between the RPE and photoreceptors, and between the RPE and Bruch's membrane, in turn leading to detachment and secondary photoreceptor degeneration [Pianta et al., 2003].

Pattern dystrophy, peripherin2

The pattern dystrophies are a group of inherited disorders of the RPE, which are characterised by bilateral symmetrical yellow-orange deposits at the macula in various distributions, including butterfly or reticular patterns. These dystrophies are often associated with a relatively good visual prognosis, although in some cases a slowly progressive loss of central vision can occur. There is usually psychophysical or electrophysiological evidence of widespread photoreceptor dysfunction [Kemp et al., 1994]. Electrophysiological findings usually reveal an abnormal pattern ERG, a normal full-field ERG, but an abnormal EOG.

Mutations in the peripherin2 gene on chromosome 6p have been identified in patients with pattern dystrophies [Nichols et al., 1993]. Peripherin was originally identified in a strain of mice with a photoreceptor degeneration known as 'retinal degeneration, slow' (RDS). The RDS mouse, which is homozygous for a null mutation in peripherin, is characterised by complete failure to develop photoreceptor outer segments, downregulation of rhodopsin expression, and photoreceptor apoptosis. Subsequently, the orthologous human peripherin gene was shown to cause autosomal dominant retinitis pigmentosa (adRP) [Kajiwara et al., 1991]. A mutation in codon 172 of peripherin has also been implicated in autosomal dominant macular dystrophy [Downes et al., 1999].

Peripherin is a membrane associated glycoprotein restricted to photoreceptor outer segment discs in a complex with the orthologous rod outer segment membrane protein 1 (ROM1). Peripherin functions as an adhesion molecule involved in the compact arrangement of outer segment discs [Travis et al., 1991]. Peripherin has also been shown to interact with the glutamic acid and proline rich region of the beta-subunit of rod cGMP-gated channels in a complex including the sodium/calcium/potassium exchanger [Poetsch et al., 2001]. This interaction may have a role in anchoring the channel-exchanger complex in the rod outer segment plasma membrane.

Dominant drusen, EFEMP1

Dominant drusen (Doyne honeycomb dystrophy) is characterised by round yellow-white deposits – drusen – under the RPE, distributed at the macula and around the optic disc, that begin to appear in early adult life. Visual acuity is maintained through the fifth decade but patients usually become legally blind by the seventh decade.

Drusen are comprised of photoreceptor outer segment products and other materials that are deposited in Bruch's membrane because they cannot be digested by the RPE cells. If enough drusen accumulate in the membrane, exchange of nutrients between the RPE cells and the adjacent choroidal blood vessels is compromised. As a result, RPE cells die, which in turn leads to photoreceptor cell. Thus in Doyne's, visual loss is usually

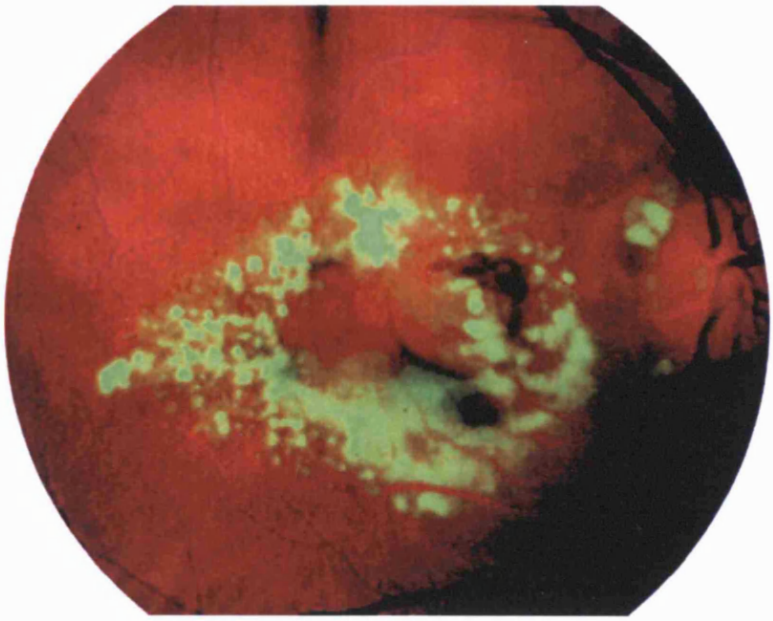


Figure 1.11: Dominant drusen. Fundus photograph showing multiple drusen-like deposits at the macula and around the optic disc. Small drusen-like deposits can also be seen to radiate from the periphery of the main drusen mass. The established complication of subretinal neovascular membrane is present centrally.

due to macular atrophy, but less commonly may follow formation of a subretinal neovascular membrane (SRNVM). The presence of drusen-like deposits makes this dystrophy potentially very relevant to age-related macular degeneration (AMD).

A single mutation (R345W) in the gene that codes for epidermal growth factor (EGF)-containing fibulin-like extracellular matrix protein (EFEMP1) on chromosome 2p has been identified in the majority of patients with dominant drusen [Stone et al., 1999]. EFEMP1 is a widely expressed gene of unknown function. Based on its sequence homology to the fibulin and fibrillin gene families, EFEMP1 is predicted to be an extracellular matrix glycoprotein. It has recently been proposed that mis-folding and aberrant accumulation of EFEMP1 within RPE cells (and between the RPE and Bruch's) may underlie drusen formation in both Doyne honeycomb dystrophy and AMD, although EFEMP1 does not appear to be a major component of the drusen [Marmorstein et al., 2002].

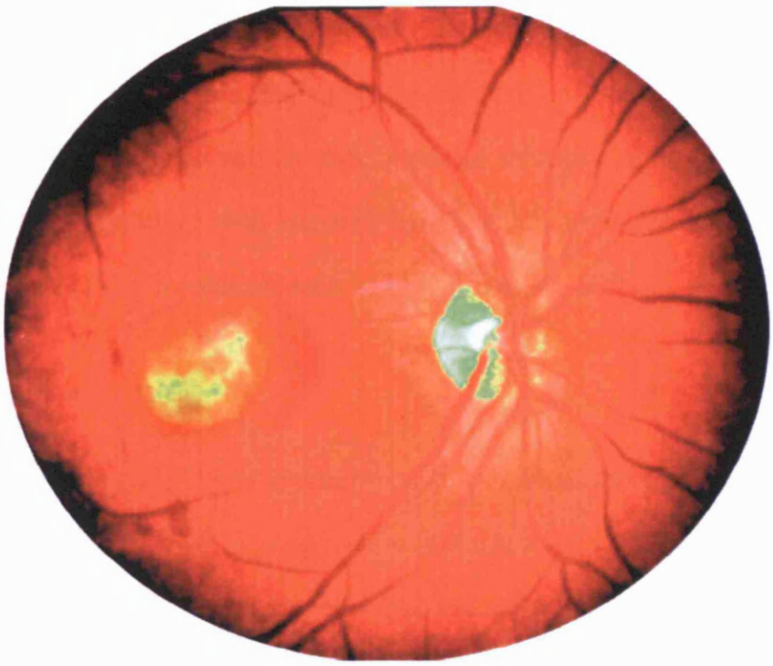


Figure 1.12: Sorsby fundus dystrophy. Fundus photography showing subretinal haemorrhage as a complication of choroidal neovascularisation.

Sorsby fundus dystrophy, TIMP3

Sorsby fundus dystrophy (SFD) is a rare autosomal dominant macular dystrophy with onset of night blindness in the third decade and loss of central vision from macular atrophy or sub-retinal neovascular membrane (SRNVM) formation by the fifth decade (see Figure 1.12).

The tissue inhibitor of metalloproteinase-3 (TIMP3) gene on chromosome 22q has been implicated in SFD [Weber et al., 1994]. Most of the known mutations in TIMP3 – including Ser181Cys, Ser156Cys, and Tyr172Cys – introduce unpaired cysteines to the C-terminus of TIMP3. This results in inappropriate disulphide bond formation and thus an abnormal tertiary structure that alters TIMP3-mediated extracellular matrix turnover. This in turn leads to thickening of Bruch's membrane and widespread accumulation of abnormal material beneath the RPE [Chong et al., 2000].

Interestingly, treatment with high doses of vitamin A reverses night blindness due to SFD [Jacobson et al., 1995]. This suggests that the retinal dysfunction may be due to reduction in the permeability of Bruch's membrane (by accumulated extracellular debris) resulting in the hindrance of transport of vitamin A from the choriocapillaris to the pho-

toreceptors. Mutant TIMP3 has also been used to induce apoptosis of RPE cells [Majid et al., 2002], suggesting that apoptosis may be the final common pathway for cell death in SFD. TIMP3 has also been shown to be a potent inhibitor of angiogenesis via its effect on vascular endothelial growth factor (VEGF), which may account for the recognised complication of SRNVM formation seen in SFD [Qi et al., 2003].

X-linked retinoschisis, RS1

X-linked retinoschisis (XLRS) is a vitreo-retinal degeneration that commonly presents in childhood with mild loss of central vision [George et al., 1995a]. The characteristic fundus abnormality is a cystic spokewheel-like maculopathy (foveal schisis) in virtually all affected males (see Figure 1.13). Peripheral retinal abnormalities – including bilateral schisis cavities, vascular closure, inner retinal sheen, and pigmentary retinopathy – are also seen in approximately 50% of cases [George et al., 1995b]. Full-field ERG typically reveals a negative waveform, in that the a-wave is larger in amplitude than the b-wave. Prognosis is good in most affected males in whom retinal detachment or vitreous haemorrhage does not occur. The histopathological findings in XLRS include splitting within the superficial layers of the retina, photoreceptor degeneration, thinning of the ganglion cell layer, and a focally absent or proliferative RPE [Manschot, 1972].

XLRS has been linked to mutations in the gene RS1 on Xp22.2 [Sauer et al., 1997]. Juvenile retinoschisis shows wide phenotypic variability both between and within families with different genotypes [Eksandh et al., 2000]. RS1 encodes a 224 amino acid protein that contains a highly conserved discoidin domain that is involved in cell–cell adhesion and cell–matrix interactions. These functions correlate well with the observed retinal schisis in XLRS.

RS1 is assembled in both photoreceptors and bipolar cells as a disulfide-linked oligomeric protein complex [Molday et al., 2001]. The secreted RS1 complex associates with the surface of these cells, where it may function as a cell adhesion protein in maintaining the integrity of the central and peripheral retina.

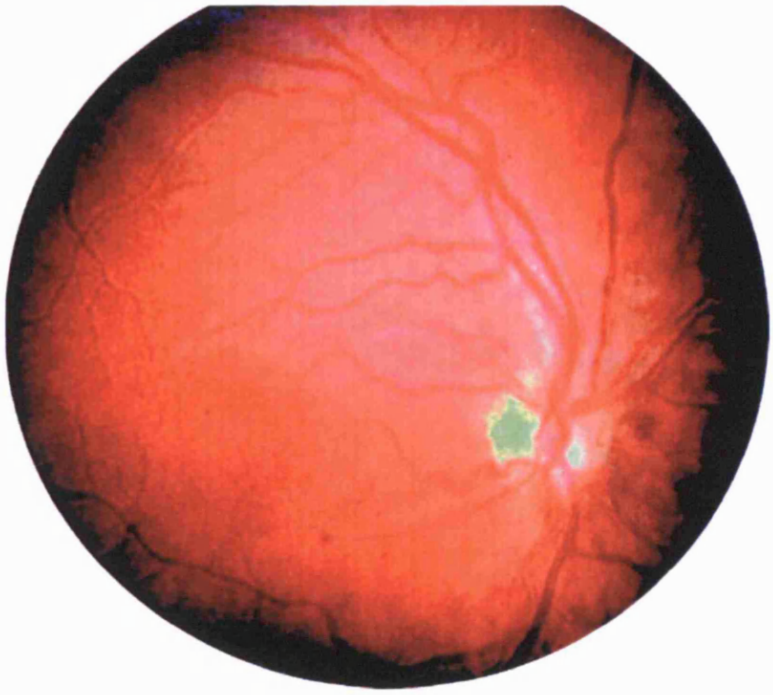


Figure 1.13: Fundus photograph showing a characteristic spoke-wheel lesion in the fovea of X-linked retinoschisis.

Knockout mice have been generated that are deficient in Rs1h – the murine orthologue of the human RS1 gene [Weber et al., 2002]. The hemizygous Rs1h-/Y male mouse was shown to share several diagnostic features with human XLRS, including the typical ERG response and the development of cystic structures within the inner retina, followed by dramatic photoreceptor cell loss. The presence of atypical ribbons at the photoreceptor terminals suggested a role for RS1 in photoreceptor-bipolar synapse formation. Failure of synaptic connections could very well lead to photoreceptor cell death.

1.3.3 Age-related macular degeneration

Age-related macular degeneration (AMD) is the most common cause of blindness in the developed world. It is now responsible for over 50% of blind and partial sighted registrations in the UK [O'Shea, 1998]. Approximately 10% of people over the age of sixty-five, and 30% over seventy-five suffer from it. Moreover, the prevalence of AMD is increasing [Evans and Wormald, 1996]. AMD causes loss of central vision.

While early AMD presents as drusen (see Figure 1.14) and related pigmentary changes,

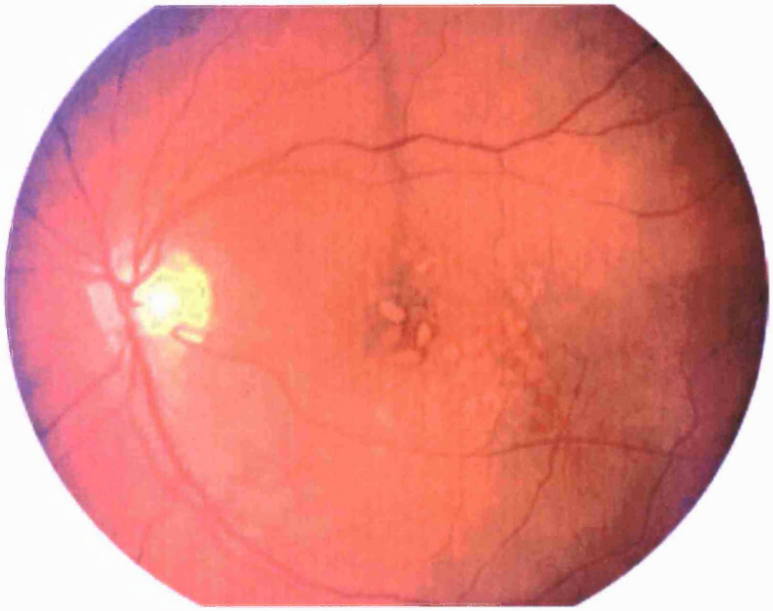


Figure 1.14: Drusen in AMD.

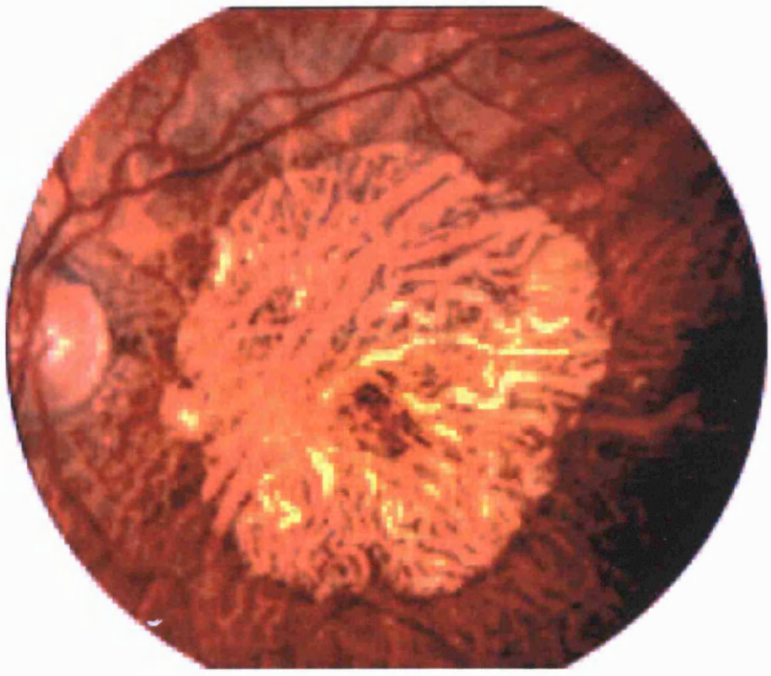


Figure 1.15: Geographic atrophy resulting in visualization of choroidal vessels underlying the retina.

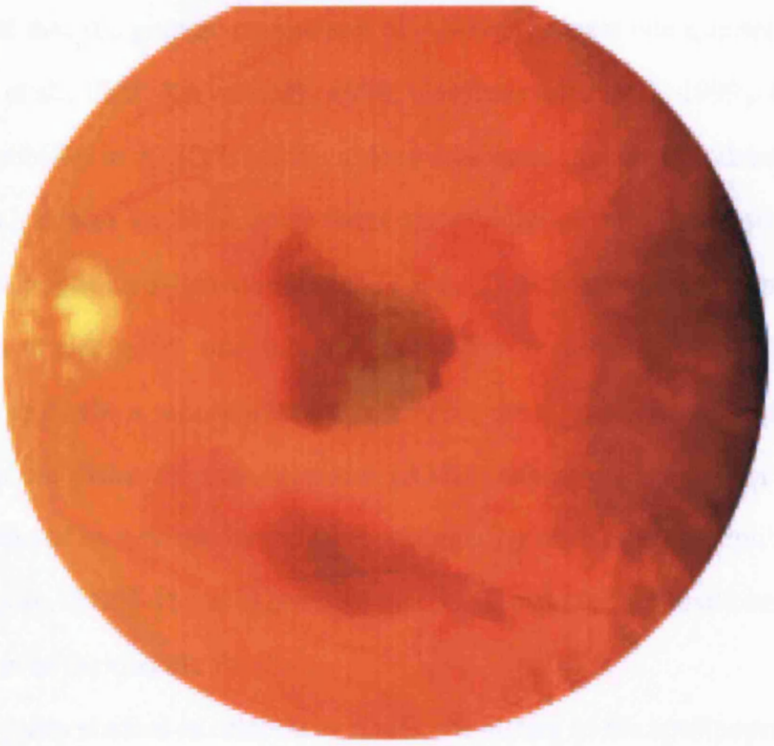


Figure 1.16: Macular haemorrhage seen in a patient with 'wet' AMD as a result of choroidal neovascular membrane formation.

late AMD can take two forms: geographic atrophy (GA, see Figure 1.14) and choroidal neovascularization (CNV, see Figure 1.16).

The pathogenesis of AMD is poorly understood and this has hampered the development of rational therapies. Options for the treatment of AMD are limited. There is some evidence that high dose vitamin C and E, beta carotene, and zinc supplementation (the AREDS supplements) leads to a lower probability of progressing from early to advanced AMD. [Group, 2001] However, the confidence interval on the odds ratio is fairly wide and barely excludes unity (0.72; 99% confidence interval, 0.52-0.98). The AREDS supplement is contraindicated in smoking – the largest known risk factor for AMD apart from age – so getting people to give up smoking may be more useful in preventing AMD than the AREDS supplement. For the treatment of GA, low vision aids and training can be offered. For CNV, the mainstay of therapy is laser photocoagulation, though only a small number of patients are suitable for treatment and recurrence is common [Yates and Moore, 2000].

Both genetic and environmental risk factors contribute to AMD. Twin and sibling

studies suggest that the genetic component of AMD is at least one quarter [Heiba et al., 1994; Meyers et al., 1995; Klaver et al., 1998; Gottfredsdottir et al., 1999]. It is likely that genetic susceptibility to AMD is attributable to 3 or more genes of moderate effect. That is, genes which confer approximately twice the population risk to the sibling (or other 1st-degree relative) of an affected individual ($\lambda_s \approx 2$). A feasible genetic model involves three subtypes of risk gene: one subtype contributes towards the overall predisposition (lifetime risk) to AMD, a second subtype influences the age of onset of AMD, and a third subtype hones the disease's characteristics (AMD phenotype, eg. GA or CNV). AMD may well be caused by environmental factors triggering disease in susceptible individuals [Yates and Moore, 2000]. It is also possible that AMD represents a final common pathway of degeneration in the macula.

Of all the genes studied in relation to AMD, the alleles of the apolipoprotein E (apoE) gene have been the most consistently associated with disease. However, not all apoE studies have found an association, and among these the associations differ. On balance, the epsilon4 allele appears to be protective, or at least delays the onset of the disease, whereas the epsilon2 allele appears to speed the onset of AMD. [Baird et al., 2004]

One approach to find genes, and thus proteins, important in the pathogenesis of AMD has been to study inherited retinal dystrophies with phenotypic similarities to AMD. Candidates for AMD have been chosen from genes in which mutations are known to cause macular disease. However, mutational analyses have failed to yield statistically significant association of sequence variants with AMD [Allikmets, 1999]. Although the ABCA4 gene has previously been linked to AMD, the story is much more complicated than first imagined as there are so many sequence variants. Allikmets [2000] gave a tally of all ABCA4 alleles as 350 to 400, making the heterogeneity of ABCA4 comparable to that of another member of the ABC superfamily – the cystic fibrosis transmembrane conductance regulator.

A second method for yielding candidate genes for AMD has been to study linkage and association of polymorphic genetic markers in families, populations, and siblings with

AMD. Unfortunately, these studies have also not revealed any significant susceptibility genes despite candidates from monogenic retinal dystrophies [Yates and Moore, 2000].

More recently, whole genome scans have been undertaken to search for genetic susceptibility to AMD [Weeks et al., 2000, 2001, 2004; Schick et al., 2003] that have yielded candidate genes from a number of loci. There is evidence – from a large AMD family and 11 unrelated individuals – that mutations in the fibulin6 gene at one of these loci (ARMD1) segregate with AMD [Schultz et al., 2003]. Very recently, mutations have been found in two genes that confer an increased susceptibility to AMD: fibulin5 [Stone et al., 2004] and clotting factor H [Edwards et al., 2005; Haines et al., 2005; Klein et al., 2005]. Clotting factor H mutations were estimated to account for about half of AMD.

Fibulin5 and fibulin6 are both similar to EFEMP1 gene, which causes dominant drusen. Fibulin6, is an adhesion molecule that contains 6 calcium-binding EGF-like domains and an Arginyl-glycyl-aspartic (RGD) cell adhesion motif. Fibulin5 is prominently expressed in developing arteries. Fibulin5 promotes adhesion of endothelial cells through interaction of integrins and the RGD motif and may play a role in vascular development and remodelling. Clotting factor H is a serum glycoprotein that controls the function of the alternative complement pathway and acts as a cofactor with factor I as a complement 3b inactivator. Factor H regulates the activity of the C3 convertases, such as C4b2a.

1.3.4 Functional genomics

An alternative approach to investigate the molecular genetics of macular degeneration is to study gene expression in the macula itself. As well as containing a higher density of cones than rods, foveal cones are known to be somewhat different from peripheral cones (at least with respect to gene expression) for example in the expression of calbindin [Haley et al., 1995]. Bernstein and Wong [1998] suggested that analyzing the regional expression of specific candidate genes could provide important clues for dissecting the pathophysiological basis for retinal disease.

Very high-titer cDNA libraries containing more than a million independent clones can

be constructed in lambda vectors, for example λTriplEx, because of the high efficiency packaging and transduction of recombinant *phage* into *E. coli* [Sambrook et al., 1989]. In λTriplEx *phagemid* vectors (Clontech) the cloning sites are within a *plasmid* that is itself embedded in a lambda phage and flanked by loxP recombination sites. Transducing λTriplEx into an appropriate E. coli strain promotes Cre recombinase-mediated release and circularization of the smaller *plasmid* pTriplEx at the loxP sites [Elledge et al., 1991].

Den Hollander et al. [1999b] isolated thirty-three cDNAs expressed specifically, or at highest levels, in the retina or RPE/choroid compared with other tissues. This was achieved by constructing a cDNA library using a polymerase chain reaction (PCR) based suppression subtractive hybridization technique that normalizes sequence abundance and achieves high enrichment for differentially expressed genes [Diatchenko et al., 1996]. A mutation in one of the genes found using this approach turned out to be the cause of retinitis pigmentosa 12 [den Hollander et al., 1999a].

As AMD susceptibility is due to both environmental and hereditary factors, it is logical to look at gene expression in order to try and understand the disease process. Indeed, our molecular genetic understanding of both the macula and AMD are likely interlinked. The macula is responsible for gathering information necessary for fine visual acuity. It is precisely this acuity – allowing us to read, write, and distinguish faces – that is lost in macular degeneration.

Functional genomics is a relatively new term that describes tissue-specific evaluation of changes in gene expression in relation to underlying differences in biology. New technologies such as DNA microarrays (see chapter 3) offer opportunities to investigate gene expression in the macula from both a functional and pathological point of view. Rather than relying on linkage of one or more genomic regions with macular disease, this approach focuses on patterns of transcript expression in retinal tissue. This may be useful to determine important candidate genes for macular degeneration. On a larger scale, functional genomics may help elicit the whole cacophony of molecular events involved in precise vision and its maintenance in the retina.

In combination with other molecular and genetic approaches, functional genomic techniques offer an invaluable tool for understanding the interactions of the many cellular processes that maintain the human macula and contribute to macular disease [Gorin et al., 1999].

1.4 Objectives

My overall objectives in studying human macular gene expression were:

1. To look for genes that are differentially expressed in macula compared with peripheral retina.

2. To characterize one or more of the proteins encoded by these genes in relation to the macula.

I aimed to characterize robust differential expression in genes not previously implicated in macular function or dysfunction.

In the course of the thesis, I will cover the specific objectives of how the human macular resource was obtained and verified, microarray analysis to find differentially expressed genes, confirmation by quantitative PCR, and characterization of histone deacetylase 9 and the *morpheus* gene family.

The first stage, towards reaching the overall objectives in studying human macular gene expression, was to obtain a source of human retinal RNA. Once access to enough fresh tissue was obtained, techniques for dissection of the macula from human donor eyes and RNA extraction using small quantities of neural retina needed to be refined. The next logical step was then to test samples of the RNA obtained for enrichment in genes that are known to be differentially expressed in the macula, such as red or green cone opsin. Finally, a cDNA library was constructed as a long-term resource.

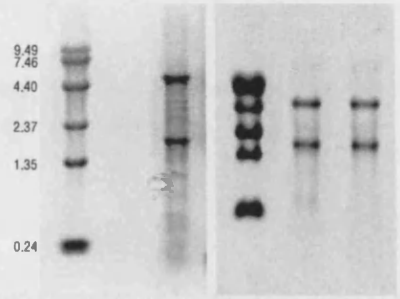


Figure 1.17: Two formamide agarose gels of retinal RNA. The left panel shows total RNA extracted from ten 2mm foveo-macular punches. RNA size markers are labelled in kilobases (kb). The right panel shows RNA from twelve pooled 2mm foveo-macular punches (left) and two 4mm macular punches (right).

1.5 Results

I procured and dissected over 60 human donor eyes and extracted RNA from over 50, using multiple sections from many eyes. For each donor eye and specimen patient details including: age, sex, times of death, morgue admission, enucleation and dissection, cause of death, medical and ocular history, and any visible ocular pathology were recorded on the consent form. This form was retained in the laboratory, which sadly suffered a fire recently.

A ten percent aliquot ($3\mu l$ of $30\mu l$) of RNA extracted using the modified TRIZOL method (2.3.2) from ten 2mm foveal punches was analysed by gel electrophoresis (Figure 1.17, left panel). Approximately 40 micrograms of total RNA was obtained ($\approx 1.3\mu g/\mu l$). Five microlitres of $1\mu g/\mu l$ RNA ladder (0.24-9.49kb) was loaded in all RNA gels to help assess the quality and quantity of the extracted RNA. Smaller aliquots ($1\mu l$ of $30\mu l$) of RNA obtained from column-based extraction of RNA from twelve 2mm punches and two 4mm punches were also run on a gel (Figure 1.17, right panel). Approximately 30 micrograms of total RNA was obtained ($\approx 0.5\mu g/\mu l$). In both gels, ribosomal 18S and 28S RNA, and several mRNA bands in between can clearly be identified.

Aliquots of RNA extracted separately from ten sections of peripheral retina from different eyes were also analysed on a gel (Figure 1.18). Thirty microlitres of total RNA was obtained from each extraction with concentrations ranging from $\approx 0.25\mu g/\mu l$ to over $5\mu g/\mu l$.

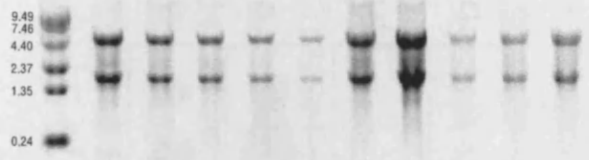


Figure 1.18: Total RNA extracted from ten individual peripheral retinal sections. RNA size markers are labelled in kb.

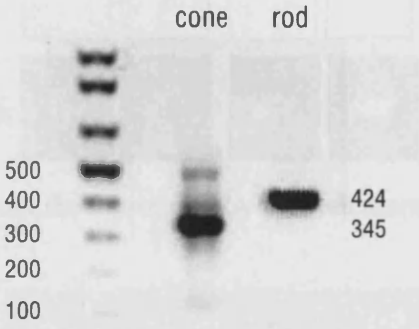


Figure 1.19: Amplification of non-homologous cone and rhodopsin cDNA from 4mm foveal and peripheral retinal cDNA respectively. The major bands were cut out before use as probes for Northern hybridization.

The PCR primer pairs for cone and rhodopsin yielded predominantly 345 and 424 base pair products, as predicted, with 1st strand cDNA from 4mm foveal and peripheral retinal RNA respectively (Figure 1.19). Faint auxiliary bands in the amplified cone opsin cDNA were likely due to the small difference in size between red and green cone opsin.

Northern analyses using the purified, ^{32}P-dCTP labelled rod and cone and opsin cDNA probes showed higher expression of cone opsin mRNA at the expected molecular weight of 1.4kb in the 4mm foveal punches compared with the peripheral retina (Figure 1.20). Conversely, higher expression of rhodopsin was seen in the peripheral retina with a band at the expected weight of 1.75kb and two additional bands. There was some difficulty loading exactly the same amount of RNA in each lane, due to the imprecise nature of quantification by gel analysis. However, cone and rhodopsin levels were higher in fovea (4mm) and peripheral retina respectively, despite less total RNA being loaded into these lanes.

Initial library analysis carried out by Clontech counted approximately 3.4 million independent clones in the *unamplified* foveo-macular library. PCR analysis of 20 ran-

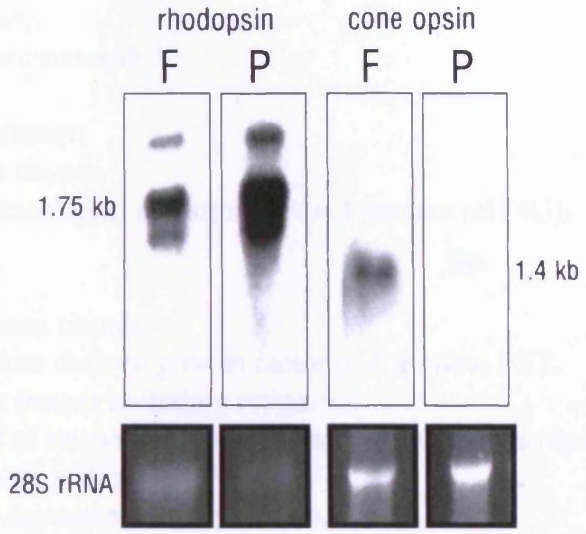


Figure 1.20: Analysis of cone and rhodopsin mRNA in fovea 4mm (F) and peripheral retina (P) by Northern hybridization.

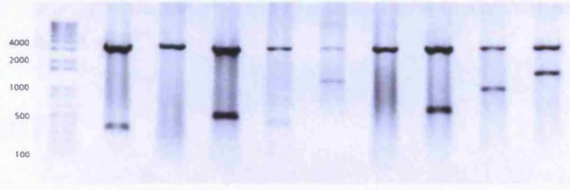


Figure 1.21: Several inserts from the foveal library. Size markers (first lane) are labelled in base pairs. Note that the upper bands represent linearized pTriplex at 3570 base pairs and that the lower bands represent the inserts. There was no apparent insert in two of the plasmids run on this gel.

domly selected clones yielded 20 unique inserts ranging between 0.5 and 7kb in length (mean 2.6kb). My initial hybridization screening revealed at least five clones in the *amplified* foveo-macular library (0.025%) using the cone opsin probe, compared with none in the *amplified* total retinal library from Jeremy Nathans' laboratory. About ten clones in each of the total retinal (2%) and foveal (0.05%) libraries lit up with a rhodopsin probe. Bi-directional sequencing and subsequent BLAST analysis of plasmid inserts (see Figure 1.21) obtained by converting the *amplified* library into plasmid revealed 11/16 non-repetitive sequences with several corresponding to genes that are known to be expressed in retina (see Table 1.1).

NADH dehydrogenase.
Repetitive DNA. Chromosome 3.
Similar to Alu.
Likely Ribosomal protein.
Human EST various tissues.
Human eukaryotic translation initiation factor 4 gamma (eIF4G).
Similar to Alu.
Similar to Alu.
Human protein, various tissues.
Human lens epithelium derived growth factor p52. Retinal EST.
Human EST various tissues including retina.
Identical to sections of known protein. Human EST. Contains repetitive DNA.
Human protein. Neural origin.
Mitochondrial ATP-dependent (lon) protease.
Similar to Alu.
Human protein, various tissues.

Table 1.1: Table of *amplified* foveal library inserts that were bi-directionally sequenced and identified by using BLAST against dbEST.

1.6 Discussion

The quality and yield of the RNA extracted from donor eyes varied from excellent to poor. Long delays between enucleation and dissection adversely affected the RNA but this did not seem like such a big factor when either the cadaver was placed quickly in the morgue or the enucleated eye was quickly refrigerated. A time between death and dissection of less than 12 hours was desirable, although not an absolute requirement to obtain good quality RNA. The quality and yield of RNA was also improved by careful attention to detail in the planning process although a speedy extraction procedure was important.

I obtained more than enough high-quality RNA, with my modifications to the TRIZOL extraction method, to construct a library using SMART cDNA synthesis. A *non-normalized, un-subtracted* library was chosen because this is the closest representation of the original source material. Normalization of the library would help if a sequencing project were undertaken to identify a large number of foveal genes as it would reduce the number of times the same common transcript was re-sequenced. A subtracted library, for example made using suppression subtractive hybridization [Diatchenko et al., 1996] and peripheral retinal material as the 'driver' that is subtracted, could be a useful fovea-specific gene

discovery tool. However, both *subtraction* and *normalization*, like any manipulation of a library including amplification, tend to diminish the abundance of rarer transcripts.

The TRIZOL extraction technique, even with my modifications, proved somewhat susceptible to carry-through of RNase, as evidenced by some of my gels that inferred increased RNA degradation after a single freeze-thaw cycle (data not shown). The two largest factors in improving RNA quality were use of the column-based RNaqueous-4-PCR kit, in which the extraction procedure is speedy, and immersing the globes in RNAlater as soon as possible after enucleation.

The nucleotide sequence for green (medium-wave) cone opsin is 96% identical to that of red (long-wave) cone-opsin and so it is very difficult to design PCR primers that will differentially amplify only one of these two cone pigment transcripts. It is almost impossible to distinguish red and green opsins by hybridization. Both red and green cone opsins are known to be expressed in the fovea, whereas blue (short-wave) cone opsin is mostly expressed in the far-periphery. The PCR for cone opsin was sub-optimal. However, as the major band from each reaction was cut and purified from the gel, it is likely that the presence of faint auxiliary bands would not have affected the cone opsin hybridization probe.

The Northern analyses show that a 4mm macular punch is relatively enriched for cone-specific genes. A single transcript was seen corresponding to red/green cone opsin as expected. Three distinct bands were seen for rhodopsin, similar to that reported by other investigators. [Bernstein et al., 1996; al Ubaidi et al., 1990] Variation in opsin mRNA size is due to the presence of cryptic polyadenylation sites in the untranslated region in the primate opsin gene. Loading exactly the same quantity of RNA sample in each lane requires a large amount for running multiple gels, especially as assessment of RNA quantity is imprecise by gel electrophoresis. Standard laboratory spectrophotometers are also imprecise in determining RNA concentration. However, because there was more RNA in the lanes that showed less expression, this was not important to achieve a good semi-quantitative result. Foveo-macular 2mm punches are likely to be even more

enriched for cone-specific, and hence other macula-specific or enriched genes. Northern blotting is an insensitive [Bustin, 2000] technique that requires a lot of RNA. It is possible to detect macula-enriched genes of lower abundance than cone opsin by using more sensitive techniques for studying differential expression such as microarrays (chapter 3) and quantitative PCR (chapter 4).

The initial screening suggested the *unamplified* foveo-macular library was of relatively high quality as there were a large number of independent clones. The inserts were also large and the vast majority differed in length. Hybridization screening using the *amplified* library demonstrated at least some enrichment for cone opsin in comparison to a widely used total retinal *amplified* library. It is likely that other macula-related genes are also enriched in this library. Direct sequencing of colonies from the *amplified* foveal library showed that the great majority of inserts contained non-repetitive DNA that corresponded to known genes or ESTs, including those represented in other retinal libraries according to dbEST. It is likely that the library amplification process resulted in the loss of lower abundance transcripts so there will be a richer transcript profile in the *unamplified* stock. Also, it is difficult to completely eliminate genomic DNA in the original library construction process, although this could have been reduced by treating the RNA with DNase. I did not treat the RNA with DNase at the time as it involved taking the RNA off ice and I was very concerned about RNA degradation due to carry-through of RNase during TRIZOL extraction.

1.7 Conclusions

I procured a resource of human macular tissue that was of high enough quality for gene expression studies. I showed that my dissection technique was accurate and that, together with RNAlater and diligent extraction, it could yield high-quality RNA that was enriched in macula-related genes. The cDNA library should also prove to be a useful long-term resource.

Chapter 2

Methods

2.1 Tissue procurement

It is not easy to procure human tissue for research purposes. This has become increasingly difficult in the UK since the Alder Hey case. In fact, it is still almost impossible to obtain human eyes for research that are in a good enough condition to obtain RNA for gene expression studies at the Institute of Ophthalmology.

I collected a human macular resource in Vancouver, British Columbia, Canada. I was based in Bob Molday's laboratory at the University of British Columbia (`www.biochem.ubc.ca/Faculty/Molday.html`). I formed a collaboration with the Eye Bank of British Columbia to gain access to human donor eyes. I produced an outline of my research aims and potential benefits to health-care and presented this both orally and in writing. I subsequently arranged for the Eye Bank to put donor eyes in RNAlater (Ambion), after removing the cornea for transplantation, in order to obviate RNA degradation. Informed consent to use the eyes for the purpose of my research was obtained from the relatives of each donor in accordance with Canadian regulations.

2.2 Dissection

I developed a technique to obtain foveo-macular punches (8mm, 4mm, 2mm) from the retinas of post-mortem donor eyes (see Figure 1.4). This involved dissecting the eyes on ice to remove the cornea (if present), iris, and lens. I placed the eye on a square of pink dental wax and used an 8mm biopsy punch to excise the macula and underlying RPE, choroid, and sclera. I carefully detached the neural retina from the RPE/choroid part of the punch so that the *macula lutea* could be visualised. I then made a second, internal punch of either 4mm or 2mm with the appropriate biopsy tool. This smaller punch was placed in a 1.5ml microcentrifuge tube. If the eye had been immersed in RNAlater, I filled the microcentrifuge with at least 5 punch-volumes of fresh RNAlater, or a minimum of 200μl for 2mm punches. Otherwise, punches were flash-frozen in liquid nitrogen. I subsequently dissected the remaining peripheral retina on ice, and preserved it in a similar manner. Flash-frozen samples were stored in a -80°C freezer. Samples in RNAlater were stored at 4°C over-night or at -80°C for longer periods.

2.3 RNA extraction

I tried three different RNA extraction techniques. Firstly, I used the TRIZOL® reagent (Gibco BRL) according to the manufacturer's protocol. I then made several modifications to this technique (see 2.3.2) and started using GlycoBlue (Gibco BRL) for pellet visualisation. However, GlycoBlue is not suitable for use with microarrays. Finally, I used Ambion's RNaqueous-4-PCR according to their protocol. It is a rapid, column-based kit and includes DNase treatment. The yield of RNA and its degradation by RNase is a critical factor when using such small amounts of scarcely available tissue.

2.3.1 Handling RNA

It is necessary to take careful precautions to avoid contaminating experiments with RNase, which is present on the surface of the skin as well as in some molecular biology reagents:

1. Use nanopure or DEPC-treated water (RNase free).
2. Wear gloves at all times. Change regularly.
3. Clean workbench and shelves.
4. Place fresh lab paper or aluminium foil on workbench.
5. Use disposable plastic-ware from unopened packages e.g. micro-centrifuge tubes, pipette tips.
6. Treat all glass, metal, and non-disposable plastic equipment with DEPC to make RNase free:
   (a) Incubate equipment in 0.1% diethyl pyrocarbonate (DEPC) at $37°C$ 1 hour-overnight.
   (b) Autoclave suitable equipment for 45 minutes on liquid cycle, otherwise rinse at least twice in RNase free water. (DEPC can methylate nucleic acids)
   (c) Dispose of DEPC appropriately (suspected carcinogen).
7. Do not freeze-thaw RNA samples unless absolutely necessary - aliquot.

2.3.2 Isolation of RNA from Small Quantities of Tissue

Preparation

1. Prepare RNase-free materials:
   - 1.5ml microcentrifuge tubes
   - metal test tube rack
   - glass mortar and pestle
   - glass homogenizer
   - 26-gauge needle and 1ml syringe

- forceps

2. Place metal test tube rack in $-80°C$ freezer for at least 2 hours.
3. Place isopropanol in $-20°C$ freezer.
4. Fill an ice bucket.
5. Label a 1.5ml microcentrifuge tube for each sample and place on ice.
6. Prepare a dish containing liquid nitrogen.

Tissue Lysis

1. Transfer microcentrifuge tubes containing tissue samples from $-80°C$ freezer to frozen test tube rack.
2. Cool glass mortar and pestle in liquid nitrogen.
3. Place mortar and pestle in frozen rack.
4. Transfer tissue samples to glass mortar with forceps.
5. Grind samples to powder with glass pestle.
6. Place mortar on ice and immediately add TRIzoL®(Gibco BRL) to begin RNA isolation.

RNA isolation

1. Keep sample on ice at all times, except where explicitly stated.
2. Add 1ml TRIzoL®(Gibco BRL) to the sample.
3. Transfer sample to 1.5ml microcentrifuge tube.
4. Vortex at high speed for 10 seconds.
5. Transfer sample to glass homogenizer.

6. Homogenize (on ice) until sample is fully dissolved in TRIzoL®(Gibco BRL).

7. Shear the genomic DNA by passing twice through a 26-gauge needle connected to a 1ml syringe.

8. Transfer the sample to a fresh 1.5ml microcentrifuge tube.

9. Add $160\mu l$ of chloroform to sample and vortex for to 30 seconds.

10. Spin at maximum speed in the microcentrifuge at $4°C$ for 5 minutes to separate the phases.

11. Transfer the top 80% of the upper (aqueous) phase to a fresh microcentrifuge tube.[1]

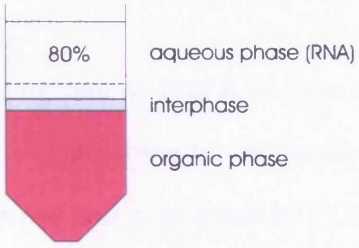


Figure 2.1: Phase separation.

12. Add $400\mu l$ ice-cold isopropanol.

13. Add $5\mu g$ GlycoBlue.

14. Allow the sample to precipitate at $-20°C$ 1 hour-overnight.

15. Pellet the RNA by centrifugation at maximum speed in the microcentrifuge for 15 minutes at $4°C$. Repeat this step until pellet is formed (blue).

16. Carefully pipette off supernatant so as not to disturb pellet.

17. Wash the pellet in $200\mu l$ of 70% ethanol and spin for 10 minutes maximum speed at $4°C$.

[1]The bottom 20% of the aqueous phase may contain protein from the interphase, including RNase.

18. Carefully pipette off supernatant, removing as much as possible without disturbing the pellet.

19. Dry briefly in a stream of nitrogen gas.

20. Dissolve the pellet in RNAse-free water. Use as much water to make a roughly $1 - 2\mu g/\mu l$ solution (use between 5 and $30\mu l$ water).

21. Vortex until pellet is fully dissolved.

22. Store at $-80°C$.

The glycogen remains in the aqueous phase and is co-precipitated with the RNA. It does not inhibit first-strand synthesis at concentrations up to 4 mg/ml and does not inhibit PCR.

I have obtained up to 3 micrograms of total RNA from a single 4mm foveal punch, and nanograms of total RNA from a single 2mm foveal punch, suitable for first strand cDNA synthesis and detection of both rod and cone opsin by PCR.

2.3.3 RNA gel electrophoresis

Materials

1. Prepare RNase-free materials:

   - electrophoresis apparatus
   - $500\mu l$ microcentrifuge tubes
   - $50ml$ measuring cylinder
   - conical flask
   - 10x MOPS solution (RNA running buffer)
   - agarose
   - RNase-free water

2. Prepare fresh RNA loading buffer (or use frozen aliquot):

   - $0.72ml$ deionized formamide
   - $0.16ml$ 10X MOPS
   - $0.26ml$ 37% formaldehyde
   - $0.2ml$ RNase-free water
   - $0.08ml$ 100% glycerol
   - $0.08ml$ bromophenol blue (saturated solution)

3. Prepare fresh dilution of 1:10 ethidium bromide with RNase free water from 10mg/ml stock solution.

Methods

1. Melt $0.5g$ agarose in $42.3ml$ RNase-free water (microwave on high for 2 minutes) in conical flask.
2. Add $5ml$ 10X MOPS.
3. When cool ($\approx$ 10 minutes) add $2.7ml$ formaldehyde (in fume hood).
4. Pour gel into taped gel dish under fume hood. Apply appropriate gel comb.
5. For each sample, mix $1\mu l$ of RNA loading buffer per μl of RNA in a $500\mu l$ micro-centrifuge tube.
6. Denature RNA by heating to $95°C$ for 2 minutes.[2] Immediately cool on ice.
7. Add $1\mu l$ of 1:10 ethidium bromide to each sample. Mix.
8. Remove gel comb from set gel. Apply samples to lanes. Run at up to $120V$.

[2]Do not microwave.

2.4 Analysis of foveal and peripheral retinal mRNA

I compared the level of rod and cone opsin expression in the 4mm foveal punches to that in the peripheral retina (*ex. fovea*) by Northern analysis. Initially, I synthesised 1st strand cDNA from total RNA extracted from a 4mm foveal punch, using reverse transcriptase (Superscript II, Gibco BRL) and an oligo-dT primer. I then designed primer pairs for rod and red/green cone opsin in order to synthesise cDNA probes for Northern analysis. The primers and probes were designed by hand and checked for specificity using the Basic Local Alignment Search Tool (BLAST) at `ncbi.nlm.nih.gov`. I then amplified rod and red/green cone opsin cDNA by using the polymerase chain reaction (PCR). This, and all subsequent PCRs were optimised by making up a standard master-mix of buffer (including magnesium), dNTP, primers, template and polymerase. In this case I used *Thermus aquaticus* (*Taq*) polymerase and varied the annealing temperature and magnesium concentration until a single strong band was seen by agarose electrophoresis. Thirty seconds extension time was allowed per kilobase (kb) of template for *Taq*.

I ran the products of the optimised PCR reactions for rod and red/green cone opsin on a gel by electrophoresis. All DNA gels were made up as ≈1% agarose in TAE buffer (40mM Tris, 20mM acetic acid, 1mM EDTA).[Sambrook et al., 1989] I purified the bands using the QiaQuick (Qiagen) Gel Extraction Kit according to the instructions. I then radio-labelled each PCR product separately with ^{32}P-dCTP by random priming (Prime-a-Gene Labelling System, Promega) before use as non-homologous cDNA probes for Northern hybridization.

I performed multiple Northern analyses for rod and red/green cone opsin mRNA levels to compare the 4mm foveal punches with peripheral retina. Firstly, I ran equal quantities of total RNA from each sample on a formamide-agarose gel (see 2.3.2 for protocol) as assessed by gel imaging. I then transferred the RNA from the gel onto a nylon membrane (Hybond N+, Amersham) by capillary transfer (see Sambrook et al. [1989]). I pre-hybridized, hybridized the membrane with the probe, and washed the membrane, using the UltraHyb™ kit (Ambion) according to the protocol. I imaged the membranes by

exposing a phosphor screen for 2-4 hours and then scanned the screen using a phosphor imaging system (Fujifilm).

2.5 Construction of cDNA library

I procured 2mm foveo-macular punches from many eyes for construction of a cDNA library. I extracted total RNA from ten of these punches. An aliquot of this RNA was analysed by gel electrophoresis. I sent off the remaining RNA for construction of a λTriplEx2 cDNA expression library. Because of the small amount of tissue available, the mRNA was first amplified using SMART^{TM}-PCR (Clontech). Whilst I did not produce the SMART cDNA for library construction, I did later perform the SMART technique myself with the Clontech kit using both foveal and peripheral retinal RNA.

2.6 Library screening

I performed a round of hybridization screening to compare the foveal library with a total retinal cDNA library, obtained from Jeremy Nathans' laboratory (`www.bs.jhmi.edu/mbg/dept/nathans.htm`), for enrichment of cone-specific transcripts. I followed the instructions in the Clontech λTriplEx cDNA library construction protocol that came with the cDNA library.

I first grew a primary streak plate – LB-tetracycline(5μ/ml) – of XL1-Blue host cells (Clontech) according to the manufacturer's protocol: a single colony from this plate was used to grow a working stock plate – LB-$MgSO_4$(20mM)-tetracycline(5μ/ml) – of host cells. From the working stock plate, I grew a single colony overnight in LB-$MgSO_4$(20mM)-maltose(20%) broth to a relative (to the growth medium) optical density (OD_{600}) of 2.0. I incubated the cells (resuspended in 20mM $MgSO_4$) with the diluted ($\approx$ 20,000 plaque forming units) library phage at 37°C for 15 minutes and plated them out in melted LB-$MgSO_4$ soft-top agarose onto an LB-$MgSO_4$-tetracycline agar(0.7%) plate. I grew the plate at 37°C overnight. I then took duplicate lifts of the plates and treated the nylon

membranes (denatured 5 minutes, neutralized 5 minutes, rinsed 30 seconds, UV light 4 minutes) to bind the DNA from the transformed bacterial colonies. Finally, I used the UltraHyb kit with ^{32}P-dCTP labelled rod and separate red/green cone opsin probes as with the Northern analyses. I imaged the membranes using the phosphor imaging system.

I also sequenced several colonies to check that they contained meaningful inserts. I first converted the library from phage to plasmid according to the λTriplEx library conversion protocol (Clontech). I transformed 200μl of BM25.8 cells (pre-incubated at an OD_{600} of 1.2 with 10mM $MgCl_2$) with 2 million plaque forming units of library lysate. I plated out the cells onto standard LB-ampicillin agar (0.7% agar, 50μ/ml ampicillin) grew them overnight and picked colonies to be grown in 5ml of standard LB-ampicillin (50μ/ml ampicillin) for a further night. [Sambrook et al., 1989] I then extracted the plasmid DNA using the QiaQuick plasmid miniprep kit (Qiagen). I performed a restriction digest using the enzyme SfiI to check for an insert and used λTriplEx2 specific primers to obtain bi-directional sequence. I checked to see what each insert contained using the BLAST with the EST database (dbEST) and the human genome database at the National Center for Biotechnology Information (`ncbi.nlm.nih.gov`).

2.7 Microarray

The microarray slides I used were designed at the Gene Array Centre, University of British Columbia (`www.prostatelab.org/arraycentre`). A set of 13,899 70-mers were obtained from the Operon (Qiagen) Human Genome Oligo Set version 1.1 (`omad.qiagen.com/human`). All probes were designed from the UniGene Database Build Hs. 147 and the Human Reference Sequence (RefSeq) Database, both developed and maintained at the National Center of Biotechnology Information (`ncbi.nlm.nih.gov`). Before choosing the array I checked that there were a substantial number of known retinal genes included on the array as controls specific to my source material.

Each microarray slide contained 28,704 spots composed of a 12×4 meta-grid of 23×26 spot grids including GFP markers at each corner and a number of control and

empty spots. Clean preparations of each 70-mer probe were spotted robotically onto the glass slides so that duplicates of each probe were in adjacent spots. The slide manufacture process was subject to strict quality-control.

I prepared the RNA targets and hybridized them to the arrays according to the Array Centre's protocols. I prepared targets with 7.5μg total RNA using either Cy3- or Cy5-dUTP and Superscript II (Invitrogen) reverse transcriptase. I cleaned up the reactions using the QiaQuick PCR purification kit (Qiagen) in light-shielded conditions and precipitated the labelled cDNA overnight at -20°C with ethanol and sodium acetate.

I resuspended each target in 50μl of hybridization solution containing: 25μl of formamide, 12.5μl of 20$\times$ SSC (3.0 M NaCl, 300 mM NaCitrate, pH 7.0), 0.5μl of 10% SDS, 5μl of 2gl^{-1} BSA, 5μl of 5gl^{-1} yeast tRNA and 2μl of 10gl^{-1} salmon testes DNA. The probe was denatured by immersing the slides in water at 95°C for two minutes just before use. The Cy3 and Cy5 target were mixed together and hybridized to the slide under a cover-slip in a moisture-rich environment at 42°C overnight.

The next day, I removed the cover-slip by floating the slide in 0.2$\times$ SSC. The slide was then washed 3 times for 5 minutes with 0.1$\times$ SSC, 0.1% SDS and 3 times for 5 minutes in 0.1$\times$ SSC. Finally, I span the slide in a falcon tube at 2000rpm for 5 minutes to dry before imaging on a Virtek ChipReader.

I overlaid grids onto images of each slide using ImageGene software (BioDiscovery), which was also used to quantify each spot in the two channels corresponding to Cy3 and Cy5 fluorescence. I parsed the raw intensity data by adapting Array Centre in-house algorithms written in the R statistical computing language (`r-project.org`) to associate each spot with its corresponding probe identity from the Operon set.

I adjusted the individual spot intensities with an Array Centre in-house background correction algorithm (see Appendix A). The algorithm is based on the observation that (due to variations in transcript levels and probe efficiencies) less than half of the probes will light up for a given target. The algorithm is used to calculate the mean of the lowest 10% of intensity values for each sub-grid as the background for that sub-grid. This

THE PROSTATE CENTRE
AT VANCOUVER GENERAL HOSPITAL
Gene Array Facility
array@interchange.ubc.ca

Protocol #S001: Processing Printed Array-It Slides

STEP 1 – Preparation of Array-It slides:

(a) Wash slides twice in 0.1% SDS for 5 minutes with shaking.
(b) Wash slides in water five times for one minute each, with shaking on a vortex at low speed.

STEP 2 - Probe Labeling and Preparation:

(a) Refer to one of the following protocols:
  (1) Protocol #S004: Direct Labeling: MIRUS
  (2) Protocol #S003: Indirect Labeling: Reverse Transcription - Cyanine Dyes
  (3) Protocol #S002: Indirect Labeling: Reverse Transcription – Amino-allyl dUTP

STEP 3 - Denaturing slides:

(a) Denature slides for 3 minutes in 95°C water bath just before use. Air dry.

STEP 4 - Hybridization of Labeled Probe to Microarray:

(a) Denature probe (prepared in STEP 2) for 3 minutes at 95°C and then keep at 65°C until use.
(b) Add probe to the denatured Array-It slide and put on cover slip.
(c) Place slide in hybridization chamber and allow probe to hybridize overnight at 42°C.

STEP 5 - Washing Hybridized Slides:

(a) Disassemble the hybridization chamber.
(b) Remove cover slip by floating the slide on 0.2X SSC.
(c) Wash the slide 3 times for 5 minutes each in Wash Solution I (0.1X SSC, 0.1% SDS) and 3 times for 5 minutes in Wash Solution II (0.1X SSC).
(d) Place slides in a falcon tube and spin at 2000 rpm for 5 minutes until dry.

THE PROSTATE CENTRE
AT VANCOUVER GENERAL HOSPITAL
Gene Array Facility
array@interchange.ubc.ca

Protocol #S003: Reverse Transcriptase Reaction with Cyanine dye-labeled dUTP

STEP 1 – Reverse Transcriptase Reaction:

(a) In the reaction, use 15-20 µg of total RNA. Set up the reaction on ice as follows:

8.0 µl	5X First Strand Buffer
1.5 µl	100 µM Anchor T primer
3.0 µl	20 mM dNTPs –dTTP (6.7 mM each)
1.0 µl	2 mM dTTP
1.0 µl	1 mM Cy dUTP
4.0 µl	0.1 M DTT
15-20 µg	RNA
to 40 µl	DEPC Water

Control RNA can also be used in the reaction (ie. Arabidopsis).

(b) Incubate the reaction at 65°C for 5 minutes, then 42°C for 5 minutes.
(c) Add 2 µl Superscript II and 1 µl RNasin.
(d) Incubate at 42°C for 2-3 hours.
(e) Add 8 µl 1 M NaOH and incubate at 65°C for 15 minutes.
(f) Add 8 µl 1 M HCl and 4 µl 1 M Tris-Cl, pH 7.5 to neutralize.

STEP 2 – Probe Purification and Denaturation:

(a) Clean up probe with Qiagen PCR purification kit or other type of spin column. Elute in the elution buffer that comes with the Qiagen kit.
(b) Now is a good time to combine Cy3 and Cy5 reactions if necessary and add 1-2 µl labeled GFP (see Protocol #S005 for GFP preparation).
(c) Optional : add 1 µl glycogen (20 µg/µl).
(d) Add 1/10 volume of 3M sodium acetate. Add 2.5 volumes of 95% ethanol. Precipitate at -20°C for at least 1 hour.
(e) Spin at top speed for 10 minutes. Wash pellet with 70% ethanol.
(f) Allow pellet to dry. Resuspend in 50 µl Hybridization solution (for one slide).

(g) For 50 µl Hybridization solution add the following:

For Oligo Slides:		For cDNA Slides:	
25 µl	Formamide	25 µl	Formamide
12.5 µl	20X SSC	12.5 µl	20X SSC
0.5 µl	10% SDS	0.5 µl	10% SDS
5 µl	2 µg/µl BSA	5 µl	5 µg/µl poly dA
5 µl	5 µg/µl yeast tRNA	5 µl	2 µg/µl BSA
2 µl	10 µg/µl salmon testes DNA	2 µl	10 µg/µl salmon testes DNA

background value is then subtracted from each spot intensity to obtain the background-corrected value. This algorithm was used in an R script that applied the correction separately for the Cy3 and Cy5 channels of each sub-grid on the array.

I then calibrated the data by writing my own R scripts (Appendix A). I used a number of standard statistical functions (see section 3.1) and specialized function libraries written in R:

1. Non-normalized log-ratios.
2. *Total intensity normalization* using the log-median as the *normalization factor* (Equation 3.5).
3. the statistics for microarray analysis (SMA) package, which includes *lowess*, *print-tip normalization* and scaling. [Yang et al., 2002b] (`stat.berkeley.edu/users/terry/Group`).
4. *Variance stabilisation* [Huber et al., 2002] (`dkfz-heidelberg.de/abt0840/whuber`).

I first compared the effect of different algorithms by producing diagnostic MA plots (see section 3.1 and Equation 3.6) of differential expression to display the spatial and intensity trends [Smyth and Speed, 2003]. I then chose the analysis method that produced the least bias in my overall data set and ran the R scripts using the parsed data from each slide. I then plotted the difference in expression (Cy3 - Cy5) against the rank of the mean expression for both the log transformed and variance stabilised data.

I mined the processed data by inserting it into a relational database using Structured Query Language (SQL). The database was implemented using mySQL (`www.mysql.com`) on a Linux (`www.linux.org`) server. By querying the database I developed a set of strict selection criteria for obtaining genes with evidence of differential expression.

2.8 Quantitative PCR

Firstly, I pooled equal volumes of 9 samples of peripheral retinal RNA from separate eyes (all but lane twelve in chapter 1 Figure 1.17) into one tube. Secondly, I measured 1.5μg aliquots of RNA from the pooled peripheral retina and foveo-macular samples by spectrophotometry. I reverse transcribed each aliquot using Ambion's RETROscript kit with random decamers according to the protocol.

Secondly, I designed primers within rhodopsin exons 3 and 4 (see Table 2.1) to amplify a product that spanned the 3rd intron (C – 166bp). I optimised these primers to amplify the corresponding 405bp genomic fragment of rhodopsin using DNA from a human control panel. I then amplified a further 1μl sample of genomic DNA, a 1μl sample of the peripheral retinal cDNA, and a 1μl sample of the cDNA at a 1:10 dilution, using the identical optimised conditions.

I generated primer pairs to amplify cDNAs from 12 genes obtained from microarray analysis chosen to undergo further study (see chapter 3 Table 3.4). Additionally, I designed primer pairs to amplify cDNAs from rod and red/green cone opsin, and four 'housekeeping' controls: 18S ribosomal RNA, acidic ribosomal phosphoprotein (ARP), β-actin (ACTB) and ubiquitin C (UBC). I designed primer pairs using the *prime* program in the GCG-Wisconsin package (`www.accelrys.com/products/gcg_wisconsin_package` to specifically amplify target cDNA using an annealing temperature of 60°C. The resultant amplicons were designed to be 100-300 base pairs long, with the region of the cDNA amplified determined by which primer pairs were most specific according to the annealing scores produced by *prime*. The additional primer-repeat and primer-template annealing tests were included to increase the specificity of the primers.

I designed the primer pairs to be used for quantitative PCR with SYBR-green PCR (Applied Biosystems) as the reporter dye. Two-step SYBR-green PCR (with a joint annealing and extension step) was performed for both primer optimisation and qPCR according to the manufacturer's protocol. I used the peripheral retinal cDNA at a 1:10 dilution to find optimal primer pairs for each target. I ran 1.5% agarose gels to check for

presence of a single clean product, otherwise I designed further primer pairs. I melted the products of reactions, that looked optimal by gel electrophoresis, in the ABI 7700 (Applied Biosystems International) to check for a single product (and to discern its melting temperature) by generating a graph of fluorescence against temperature. To double-check the specificity of the primers, I directly sequenced 6 of the PCR products. Optimal primer pairs are shown in Table 2.1.

I ran quantitative PCR reactions for each target to compare amplification from the peripheral and foveo-macular samples. I diluted a fresh aliquot of each cDNA sample 1:10 and used the same diluted aliquot for all reactions, except the 18S rRNA, in which both aliquots had to be diluted a further 1:100. Each target was amplified in triplicate from foveo-macula, peripheral retina, and a no-template (water) control. An additional PCR step was inserted in the ABI 7700 program at an appropriate melting temperature at which to measure the fluorescence of the product. This was determined from the melting curves. I obtained fluorescence values for each cycle using the ABI software that came with the 7700 and exported them as a spreadsheet.

I analysed the data using the Data Analysis for Real-Time PCR (DART-PCR) system. DART comprises several stages of data analysis that are contained within an Excel (Microsoft) workbook (see Appendix B). DART uses the amplification plot method (see section 4.1) to determine the reaction efficiency (E) and threshold values ($f(c_t)$). I used the mean efficiency and threshold for the two triplicate reactions (foveo-macula and peripheral retina) to determine the initial fluorescence values (f(0)) for each primer pair.

On a separate spreadsheet, I calculated normalized f(0) values for each test gene by dividing the calculated f(0) by the geometric mean of the the 4 control gene f(0) values. I calculated normalized f(0) values for foveo-macula and peripheral retina separately and then obtained the relative expression ratio in foveo-macula:peripheral retina. I then computed the $\log_2$expression ratios and plotted these on a graph.

After normalization to the 4 control genes, I determined the expression of each test gene as a percentage of opsin expression. I took only the genes which were two or more

Symbol	Accession	Start	Primer pair	End	Total	Amp
RHO	NM_000539	662	CGACTACTACACGCTCAAGCCGGAGG			
			CGTAGGGCACCCAGCAGATCAGGAAAGC	900	2767	239
RHO	NM_000539	1394	CCCTCCTCCTCCCAACTCATCTTTCAGGAACAC			
			GCCCAGGGAGGGAAAAACAACTGCTCCATATC	1643	2767	250
OPN1LW	NM_020061	923	CCTACTTTGCCAAAAGTGCCACTATCTACAACCCC			
			TCATGCAGGCGATACCGAGGACACAGATGAG	1095	1095	173
6q	AL133101	1931	ACAGAGAGAACCTTTTGTGCCTTTTTGCCTGC			
			CAGGGAAGGCTGGGCTCCAAAGACATCTATC	2155	2715	225
BHM2	NM_017614	1152	GAAAAAGTTGCCCTCAAGCCTGACCTGGAAC			
			CAAAGGAATGCACCACAGCATGATTGGACAGTAAG	1401	2007	250
CRP	NM_016280	137	GGCAAAGTGTTGGGGAAGTTCATCAGCTTAGAAGG			
			AAATAGCTCTGAGATTAACTGCCCCCCCCGC	352	2270	216
FLJ	AK000110	41	CCAGGATAAAAACAGACCGTGTCTCAGTAACTGGC			
			ACAAAAGAGCACTGTTCCCAACTGATGACACAGG	195	1798	155
GCN2	AB037759	1160	CCCAGCCAAAAATGCCTCTAGTGGAACAAAGTCC			
			CTGCACCTTGATGACAGCTCCAAAAGCTCC	1362	4994	203
HDAC9	NM_178423	3085	CCACTTGCAGAAGATATTCTCCACCAAAGCCCG			
			ACACAGGAAATATCAGAGGGGGACTTGGCAC	3383	4659	299
NPIP	NM_006985	452	GCATGAAAGAACGTGAGCACGGAGAAAAGGAGAG			
			TTTTGCATCTGTAGTAGTCCTCTGCCCGCC	607	1070	156
RP2	NM_006915	845	GGGGTCAGAGACAGAAGAGCAGCGATGAATC			
			ACCATCCCCATTAAACTCCAAGGCAATAACAGGAC	1090	3813	246
SAG	X12453	816	CAGTTCTTCATGTCTGACAAGCCCCTGCAC			
			GCACTTTTTCTTGCGCTTCCTCCATAGCCAC	1038	1582	223
SRGAP2	AB032982	879	AACATTGCCAAGAGGAGAGCCAACCAGC			
			TCCCACCTCGGGCAGCAGATGAAAGAAAAC	1065	5677	187
X	AL110203	80	TCCCAAGGGAATACTGCAACATGCAGGCAC			
			TCGGCAGTCTTCTGATCCTGTCAAGCAGAAAC	240	2060	161
YAP65	X80507	1056	AAGCATGAGCAGCTACAGTGTCCCTCGAACC			
			TCTCCTTCCAGTGTTCCAAGGTCCACATTTGTC	1200	5153	145
18S	X03205	1233	GCGGCTTAATTTGACTCAACACGGGAAACCTCAC			
			TCGTTATCGGAATTAACCAGACAAATCGCTCCACC	1382	1869	150
ARP	M17885	590	AGCCACGCTGCTGAACATGCTCAACATCTC			
			TCCACAGACAAGGCCAGGACTCGTTTGTACC	838	1097	249
ACTB	NM_001101	1348	CTGGAACGGTGAAGGTGACA			
			AAGGGACTTCCTGTAACAATGCA	1487	1793	140
UBC	NM_021009	20	ATTTGGGTCGCGGTTCTTG			
			TGCCTTGACATTCTCGATGGT	152	2192	133

Table 2.1: PCR primer pairs. The first pair was used to amplify exons 3 and 4 of rhodopsin to assess the cDNA. The remaining pairs were the final set used for quantitative PCR. *Accession* numbers for the source sequence are given. The *start* and *end* co-ordinates of the amplified region, as well as the *total* length of the source sequence and amplicon length (*amp*) are also shown.

fold over- or under-expressed. I then divided the f(0) value in the appropriate part of the retina by the f(0) value for the corresponding opsin in that part of the retina. For example, for s-antigen (rod arrestin) I divided its f(0) value in the peripheral retina (where it was over-expressed) by the f(0) value of rhodopsin in the peripheral retina. I multiplied this value by 100 to determine that s-antigen was expressed at approximately 31% the level of rhodopsin in the peripheral retina. Conversely, for GCN2 I divided its f(0) value in the *foveo-macula* by the f(0) value for cone opsin in the foveo-macula. I then corrected these values for amplicon length (Table 2.1) by multiplying the percentage by the amplicon length ratio of test gene:opsin.

For genes of interest, I performed several bioinformatic analyses and looked at previously published data. I obtained information about gene structures by using tools at `ncbi.nlm.nih.gov`, including the UniGene, Online Mendelian Inheritance in Man (OMIM), Gene Expression Omnibus (GEO) databases, and the Basic Local Alignment Search Tool (BLAST). I also used the Ensembl genome browser (`www.ensembl.org`), and the GenomeScan (`genes.mit.edu/genomescan`), GenScan (`genes.mit.edu/GENSCAN`) and est2genome (part of the EMBOSS suite `www.emboss.org`) gene prediction tools.

2.9 Tissue and antibody resources

Bob Molday's laboratory very kindly sent me cryo-sections of human retina. I was also very grateful to receive a rabbit polyclonal antibody to the C-terminus of HDAC9 (NP_848512, 1069 amino acids) by Arthur Zelent's laboratory at The Institute of Cancer Research (`www.icr.ca.uk`). The antibody was made to the peptide:

DVEQPFAQEDSRTAG

from amino acid 1046-1060 and is specific to the full-length protein. They also sent me a human B cell precursor leukemia (REH) cell line, in which full-length HDAC9 is highly expressed.

2.10 Tissue culture

I maintained the REH cell line as a control for the expression of HDAC9 protein. REH is a line of suspension cells although I have also been responsible for maintaining adherent cell lines – such as COS (simian fibroblast), 293T (human renal epithelial) and SK-N-SH (human neuroblastoma) – at various times during my research. Tissue culture is a complex discipline. I learned the techniques applicable to my research from expert teaching, observation, listening, note-taking and periods of supervision in the early stages of my own culture experiments. A large amount of information is also available in many books and online, for example from the handbook produced jointly by Sigma (`www.sigmaalrich.com`) and the European Collection of Cell Cultures (ECACC, `www.ecacc.org.uk`).

I grew the REH cells in a 250ml culture flask at 37°C in 15ml of complete culture medium, which consists of: RPMI (Gibco, for REH suspension cells)/DMEM (Gibco, for most adherent cell lines), 10% (v/v) fetal bovine serum (FBS), 2mM L-Glutamine, 1% (w/v) Penicillin and 1% (w/v) Streptomycin. I passaged the cells twice per week according to the following protocol:

1. Warm complete medium to 37°C in a water bath.
2. Check cells under microscope for confluence.
3. Mix the cells thoroughly in the medium by pipetting up and down.
4. Draw all of the cells into the pipette.
5. Pipette 4-5ml of the cells back into the flask.
6. Pipette the remaining suspension into a centrifuge tube.
7. Add fresh medium to the flask up to 15ml.
8. Incubate the flask at 37°C.
9. Centrifuge the remaining suspension at low speed for 5 minutes.

10. Discard the supernatant.
11. Re-suspend in 15ml PBS (2.7 mM KCl,4.3 mM Sodium Phosphate Dibasic (Na_2HPO_4), 1.8 mM Potassium Phosphate Monobasic (KH_2PO_4) 137 mM NaCl, pH 7.2).
12. Repeat last 3 steps another two times.
13. Spin down once more.
14. Discard the supernatant.
15. Freeze the cell pellet at -80°C.

I saved the cell pellet for Western blotting. Procedures for adherent cells are very similar except that they must be detached by gentle trypsinisation for 1-2 minutes prior to splitting:

1. Draw off all supernatant with a pipette. Take care not to disturb the cell layer.
2. Wash once, gently, with PBS to remove any debris.
3. Pipette off PBS.
4. Add just enough trypsin to cover the cell layer ($\approx$0.5mL in 25ml flask).
5. Incubate for 2 minutes maximum at room temperature.
6. Shake and tap occasionally to verify that the cells are releasing.
7. Add 10mL of complete medium.
8. Mix the cells thoroughly by pipetting up and down until there are few clumps.
9. Draw all of the cell suspension into the pipette.
10. Leave 1-2ml in the flask.
11. Discard the rest or save for an experiment.
12. Add 25mL of fresh medium.

13. Incubate flask on its side at 37°C.

To prepare a cell extract from the frozen pellet I re-suspended it in ≈ 5ml of PBS and passed it twice through a 26-gauge needle to shear the genomic DNA. To prepare a retinal extract, I homogenized retinal tissue in the smallest possible volume of PBS and also passed it twice trough a needle – I obtained a small amount of donor tissue from Pathology at the Institute of Ophthalmology.

2.11 Enzyme linked immuno-sorbent assay (ELISA)

2.11.1 Coating

1. Add 100ul peptide at 5-10μg/ml in 50mM carbonate buffer ≈pH 9.5:

Na_2CO_3	1.59g
$NaHCO_3$	2.93g
NaN_3	0.20g

2. Make up to one liter with distilled water.

3. Incubate at room temperature overnight in a humid chamber.

2.11.2 Blocking

1. Remove coating solution, rinse twice with distilled water, and flip dry on paper towel.

2. Add 200μl 0.1% BSA-PBS-0.02% thimerosal.

3. Incubate for one to two hours at room temperature in a humid chamber.

2.11.3 Sample

1. Remove blocking solution, rinse twice with distilled water, and flip dry on paper towel.

2. Prepare serum dilutions by diluting in the blocking buffer.

3. Prepare the following dilutions (make 1:5 serial dilutions):

Pre-immune	Test bleed
1:1000	1:1000
1:5000	1:5000
1:25000	1:25000
	1:125,000
	1:625,000

4. Add 100ul of serum samples to the wells. Run in duplicate.

5. Incubate for one to two hours at room temperature in a humid chamber.

2.11.4 Wash

1. Dump out the serum dilutions and wash the wells thoroughly and vigorously five times with PBS-0.02-% thimerosal-0.05% tween.

2. Wash five times as above with distilled water.

3. Flip dry.

2.11.5 Conjugate

1. To each well, add 100μl of Goat anti rabbit-HRP conjugate (similar to Boehringer 605 220) 1:10,000 dilution in 0.1% BSA-PBS-0.02% thimerosal.

2. Incubate at room temperature for two hours in humid chamber.

2.11.6 Wash

1. Dump out the conjugate and wash the wells thoroughly and vigorously with PBS-0.02% thimerosal-0.05% tween, three times.

2. Wash three times with distilled water. Flip dry.

2.11.7 Substrate

1. Add 100ul of 3,3',5,5'-Tetramethylbenzidine (TMB) soluble substrate and develop at room temperature for ten minutes.

2. If the color develops too quickly, dilute the peptide further.

2.11.8 Stop

1. Add 50μl 1N HCl to each well to stop the color development

2. Read immediately at A450.

2.12 Western blotting

I used Western blotting to study HDAC9 protein expression in cell and human retinal extracts. Western blotting is a widely used method of *in vitro* protein detection that entails electrophoretic separation and immobilization of protein followed by antibody binding and imaging.

I separated protein using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). I prepared 10% gels, according to a standard recipe [Sambrook et al., 1989]. To the extracts I added 5$\times$ sample buffer (SB) containing 62.5mM Tris-HCL pH 6.8, 2% (v/v) SDS, 10% (v/v) glycerol, 0.1% (w/v) bromophenol blue and 5% (v/v) β-mercaptoethanol, to chemically reduce the proteins in the extracts and boiled for 2 minutes for denaturation. I ran gels at 200V until the dye front just reached the end.

I then transferred the protein onto a nitrocellulose membrane using a semi-dry electrical transfer unit (BioRad). I pre-soaked the membrane in water and 6 pieces of the same size filter paper (Whatmann) in transfer buffer (TB, 25mM Tris, 192mM Glycine, 20% MeOH, 0.1% SDS) for 15 minutes. I attached the protein gel to the top of the membrane

and sandwiched both above and beneath 3 filter papers. I transferred the protein at 20V, limited to 0.4A per membrane for 20 minutes.

I then processed the membrane containing immobilised protein according to the following protocol. All incubation steps were carried out at room temperature on a rocker. Alternatively, the blocking[3] step was allowed to proceed in the refrigerator over-night:

1. Block in PBS-5% milk-0.05% Tween (5% Milk) for 1 hour.
2. Incubate with 1:500 primary antibody in 5% Milk for 1 hour.
3. Wash 5× for 5-10 minutes in PBS-Tween.
4. Incubate in 1:30,000 secondary antibody (goat anti-rabbit horseradish peroxidase) in 5% Milk for 1 hour.
5. Wash 4× for 5-10 minutes in PBS-Tween.
6. Wash for 5-10 minutes in PBS.
7. Perform enhanced chemiluminescence.

Enhanced chemiluminescence (ECL) was performed by placing the membrane on a flat surface and adding the mixed ECL reagent from the kit (Amersham) so that it just covered the blot for 5 minutes. Excess reagent was allowed to drip from the blot and it was placed face down on a fresh piece of cling-film. The blot was sealed by folding the cling-film over the top of it. It was then imaged by exposing it to X-ray film inside a light-tight cassette for 10 seconds to several minutes in the dark-room. The film was developed immediately to examine the signal and adjust exposure accordingly.

I scanned films in un-compressed tagged image file format (TIFF) at 300-600 dots per inch (dpi) using a flat-bed scanner (HP). I adjusted levels, brightness and contrast of the *whole* image in Photoshop (Adobe) *before* cropping any un-informative parts of the

[3]Nucleic acid hybridization requires a mobile source of nucleic acid (such as salmon testes DNA or yeast tRNA) to prevent non-specific binding of target (microarray) or probe (Northern blotting) to the support. Similarly, detection of protein using antibodies requires the use of mobile proteins such as those in milk (Western blotting) or bovine serum albumin (immunohistochemistry).

image. I imported images to Illustrator (Adobe) to add labels and text and saved them as (encapsulated) postscript.[4]

2.13 Immunofluorescence

2.13.1 Cell culture

I used 8-well chamber *glass*[5] slides (Nunc) for immunofluorescence, which require a minimum volume of 100μl to cover the slide surface within each chamber. I chose to use SK-N-SH adherent cells[6] for immunofluorescence studies. I used bovine serum albumin (BSA) against non-specific binding of the primary and normal donkey serum (NDS) against the secondary antibody. All incubation steps were carried out at room temperature and washes involved carefully pipetting the solution into each chamber and then carefully pipetting it out of each chamber with a pasteur pipette. I used the fluorescent die 4',6-Diamidino-2-phenylindole (DAPI), which forms fluorescent complexes with double-stranded DNA to stain nuclei:

1. Wash 2× with PBS.
2. Fix in PBS-3.7% paraformaldehyde 15 minutes.
3. Wash 2× with PBS.
4. Permeabilize with PBS-0.1% Triton 15 minutes.
5. Wash 2× with PBS.
6. Block with 100μl PBS-3% BSA-10% NDS per well 1 hour.

[4]TIFF is a raster (bitmap) file format. Raster files (such as TIFF, JPEG, or BMP) are used to store colour (including grey-scale) information about each pixel in an image. Vectors are paths such as lines and arcs, and include text fonts. Raster files cannot store vector information in a scalable format. Postscript files can contain both raster and vector information. Portable document format (PDF) is Adobe's distillation of a postscript file.

[5]Plastic slides can be a big problem with confocal microscopy, causing the platform to jump minutely because of electrostatic charge build-up.

[6]I previously used REH cells to try and detect HDAC9 by immunohistochemistry by spinning down 50-100μl of cells onto a slide using a Cytospin® (Shandon) centrifuge but cell morphology was poor using this technique with suspension cells.

7. Incubate with 200μl 1:250 primary antibody in 3% PBS-BSA per well 1 hour.
8. Wash 5× with PBS.
9. Incubate with 100μl 1:100 Cy5 donkey anti-rabbit secondary antibody in 3% BSA-PBS per well.
10. Wash 2× with PBS.
11. Wash with 1:5,000 DAPI in PBS for 5 minutes.
12. Wash 2× with PBS.
13. Remove all PBS. Remove chambers from slide.
14. Add 1 drop of DAKO fluorescent mounting medium.
15. Mount with an appropriately sized coverslip.

2.13.2 Retinal sections

The retinal cryo-sections that Bob Molday's lab sent me were on standard glass microscope slides; I stored them at -80°C. I processed them under as similar conditions as possible to those used for immunofluorescence. Incubation and washing times were lengthened because of the increased time for penetration of the tissue sections. A vertical slide holder was used for washes and a horizontal moisture chamber for incubations:

1. Allow the cryo-sections to thaw and air-dry for 10-20 minutes.
2. Wash in PBS for 20 minutes.
3. Fix in PBS-3.7% paraformaldehyde for 15 minutes.
4. Wash 2× with PBS for 20 minutes.
5. Carefully let the excess PBS drain onto a paper towel.
6. Place the slide in a moisture chamber.

7. Pipette ≈200μl of PBS-3% BSA-10% NDS onto the slide to cover all sections whilst still retaining surface tension.
8. Block for 1 hour at room temperature.
9. Let the blocking solution run off the slide onto a paper towel.
10. Pipette ≈200μl 1:250 primary antibody in 3% PBS-BSA onto the slide.
11. Incubate over-night at 4°C in the moisture chamber.
12. Wash 2× with PBS for 30 minutes, drain.
13. Pipette ≈200μl 1:50 Cy5 donkey anti-rabbit secondary antibody in 3% BSA-PBS onto the sections with the slide in a moisture chamber.
14. Incubate for 4 hours at room temperature.
15. Wash 2× with PBS for 30 minutes.
16. Wash with 1:5,000 DAPI in PBS for 30 minutes.
17. Wash 2× with PBS for 20 minutes, drain.
18. Add 1 drop of DAKO fluorescent mounting medium.
19. Mount with an appropriately sized coverslip.

2.13.3 Imaging

I initially checked slides on a fluorescence microscope and subsequently acquired images on a confocal microscope. The confocal is an Axioplan (Zeiss) laser scanning microscope. I used three different lasers – an Argon laser to excite Cy5 and for simultaneous differential interference contrast (DIC, Nomarski), a Helium-Neon laser to excite Cy3, and an ultraviolet laser to excite DAPI. I firstly gained a general overview of the sections on the slide at low power and then changed to a higher power to find representative cells or

illustrative regions of tissue. I then acquired several images using the software provided with the integral camera system. I used the same software to export the highest quality images (and scale bars) as TIFFs. I adjusted the levels, brightness and contrast of *whole* images in Photoshop (Adobe), assembled multiple images and added labels and text in Illustrator (Adobe), and saved as (encapsulated) postscript.

2.14 Antisera production

Firstly I performed bioinformatic analyses, with the help of the GCG-Wisconsin package, in order to design a peptide to which an antibody could be raised for detection of NPIP protein. I obtained the 70 base pair sequence of the oligonucleotide microarray probe for NPIP from Qiagen.

```
 1 CTCCCAAGTG TGTCTGCTCA CTCCCCTTCC
31 ACCCTCAGCG GATGATAATC TCAAGAAACT
61 AAGGAAGAAT
```

I then aligned both the 70mer and the PCR primer pair used for quantitative PCR (see chapter 4 Table 2.1) with the reference sequence for NPIP (AF132984). I also lined up possible splice variants from ESTs and gene prediction programs (see chapter 4 section 2.8). I designed a peptide that was represented in known and predicted isoforms of NPIP and that would likely be antigenic and not within a transmembrane domain.

I sent the 15 C-terminal amino acid sequence:

CSLPFHPQRMIISRN

of NPIP to Sigma (www.sigma-genosys.eu.com) for synthesis of 10mg peptide and subsequent production of two rabbit polyclonal antisera. After synthesis it was analysed by mass spectrometry and high performance liquid chromatography (HPLC). I requested 2-3mg of the peptide to be thiol cross-linked via the N-terminal cysteine residue, using the hetero-biofunctional (NH2,SH) m-Maleimidobenzoyl-N-hydroxysuccinimide

ester (MBS), to the carrier protein keyhole limpet hemacyanin (KLH)[7]. The remaining unconjugated peptide was sent to me. Although I did not personally carry out the immunisation protocol (see Table 2.2) on this occasion, I previously oversaw a very similar procedure for a side-project in Canada. Institutional Animal Care and Use Committee (IACUC) guidelines were followed.

Day	**Procedure**
1	Pre-bleed (15ml), antigen injection (200μg)
14	Antigen injection (100μg)
28	Antigen injection (100μg)
35	Bleed 1 (15ml)
42	Antigen injection (100μg)
49	Bleed 2 (15ml)
56	Antigen injection (100μg)
63	Bleed 3 (15ml)
70	Antigen injection (100μg)
77	Bleed 4 (50ml)

Table 2.2: Immunisation/bleeding timetable. The conjugated peptide was injected subcutaneously at multiple sites. The initial immunisation was given in Complete Freund's Adjuvant with all subsequent immunisations given in Incomplete Freund's Adjuvant.

Two New Zealand White rabbits were immunised with the peptide according to the timetable (Table 2.2). The rabbits were around 10 weeks old at the start of the protocol and have a normal life expectancy of 5-6 years. The KLH-conjugated peptide was dissolved in PBS at a concentration of 1mg/ml. For each injection, 100μg of this antigen was diluted in 1ml sterile saline and mixed thoroughly with 1ml of the adjuvant to form an emulsion. The injections were given subcutaneously at different sites every 2 weeks. Bleeds were taken through the central ear artery with a 19 gauge needle. The blood was allowed to clot and retract at 37°C for 12 hours. It was then refrigerated for a further 12 hours before the serum was decanted and clarified by centrifugation at 2500 RPM for 20 minutes.

After receiving the final bleed, I checked that both polyclonal antisera bound to the NPIP C-terminal peptide by enzyme-linked immunosorbent assay (ELISA) using horseradish peroxidase (HRP) in conjunction with antiserum bound to a positively charged micro-well

[7]KLH is typically preferred over bovine serum albumin (BSA) due to the higher immunogenicity of KLH. BSA is often used as a blocking agent in many experimental assays. Because antisera raised against peptides conjugated to BSA will also contain antibodies to BSA, false positives may occur.

Name	Sequence	T_m
EcoRI-NPIP-2	GAATTCatgttttgttgcttaggatatgaatg	55
EcoRI-NPIP-3	GAATTCatgttttgttgcttaggatatgaatggc	58
SalI-NPIP-2	GTCGACttatttattcttccttagtttcttgag	55
SalI-NPIP-3	GTCGAC.....cttagtttcttgagattatcatccgc	59

Table 2.3: PCR primers used in cloning of full-length NPIP cDNA. Restriction sites are shown in capitals. Dots are inserted to show alignment.

plate to which I had bound the NPIP C-terminal peptide. For details of the protocol see 2.3.2. I initially tried to characterize the antisera by Western blotting of extracts from COS (simian fibroblast), 293T (human renal epithelial) and SK-N-SH (human neuroblastoma) primate cell lines. I also used the antisera for immunofluorescence studies in SK-N-SH cells (see chapter 5) and in human foveal cryo-sections kindly provided by Bob Molday's laboratory in Vancouver.

2.15 Cloning

Because no previous work had been done to investigate the expression of NPIP protein, I decided to design a construct that contained a full-length NPIP cDNA obtained from neural tissue (brain or retina). For an overview of the cloning strategy I adopted, please see Figure 2.2.

First, I designed primers (Table 2.3) to add restriction sites to each end of NPIP so that it could be sub-cloned into expression vectors. A 5' *EcoRI* site and 3' *SalI* site were chosen as they are compatible with in-frame sub-cloning into a glutathione S-transferase (GST) fusion vector pGEX-6P-1 (Amersham), a maltose binding protein (MBP) vector pMAL-c2X (New England Biolabs), and also a GFP fusion vector pEGFP-C2 (Clontech) that fuses the insert to the C-terminus[8] of GFP.

Two vectors (pGEX and pMAL) were chosen for prokaryotic expression of soluble NPIP as they are suitable for use with a two-vector strategy[9] for future monoclonal anti-

[8]I find Clontech's nomenclature confusing here as the resultant fusion protein will have GFP at its N-terminus.

[9]Mice can be immunized with one fusion protein and serum and subsequent clones are screened with the other so as not to pick antibodies against the fusion protein. This can be more successful than trying to

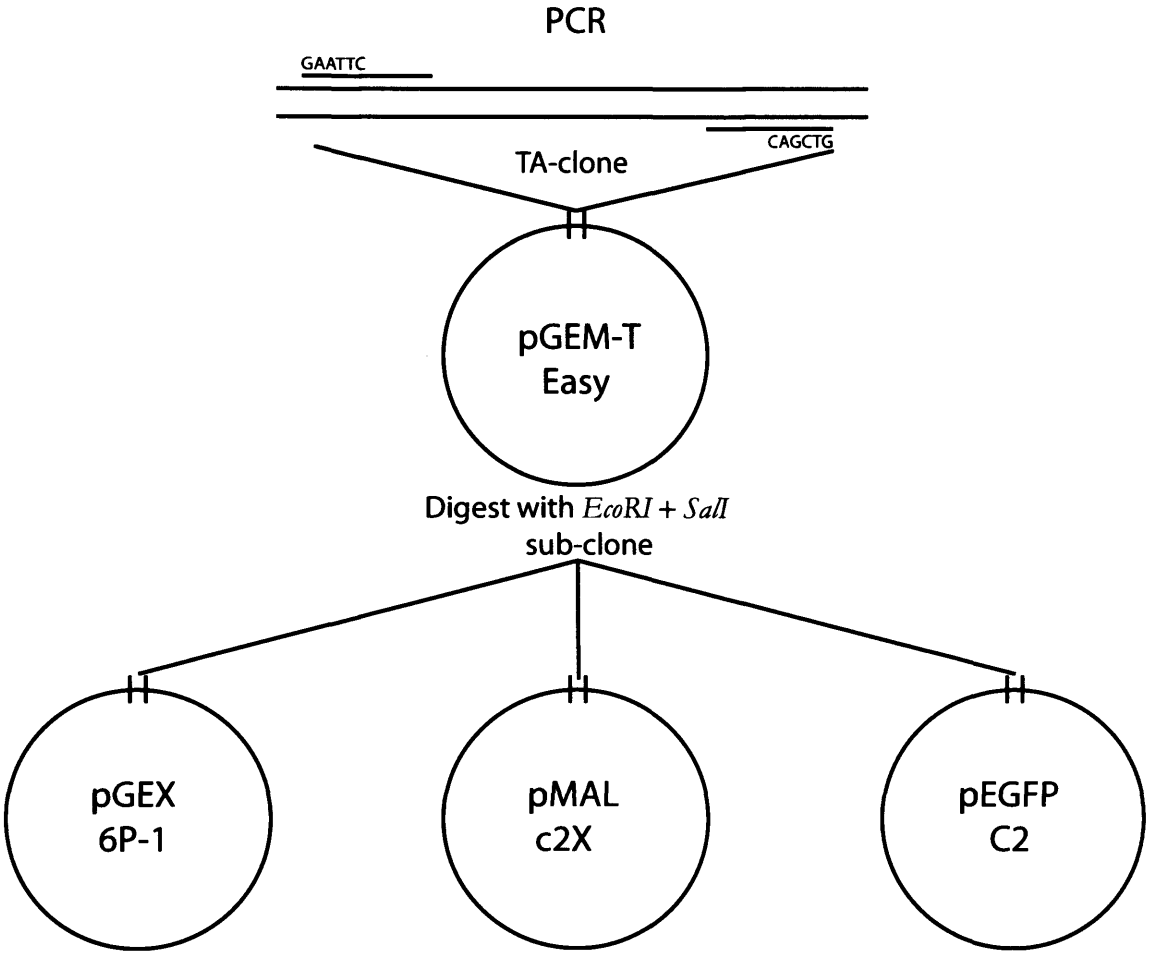


Figure 2.2: Schematic diagram to show the cloning strategy adopted. The inserts were checked for size on a gel after digestion of pGEM-T Easy and for integrity by direct sequencing before sub-cloning.

body production (as I have carried out for a side-project in Canada).

Initially, I amplified NPIP from brain and retinal cDNA (from a human brain and my foveal library), gel purified the products and cloned them into the vector pGEM-T® Easy (Promega). I used the high-fidelity enzyme with proofreading activity *Pfu*[10] from New England Biolabs (NEB) to amplify the products and then added 0.5μl of *Taq* (Bioline) polymerase and ran an extra cycle before gel purification as *Pfu* does not add the A-tail necessary for TA-cloning.

I estimated the amount of insert to ligate into the pGEM-T Easy vector from the intensity of the band on the gel before purification and by following the guidelines in the manufacturer's protocol. I used a ratio of insert:vector between 2:1 and 5:1 by volume. I ligated the insert into the vector by mixing a reaction using a concentrated T4 ligase (NEB) as follows:

10× buffer	1μl
pGEM-T Easy	1μl
Insert DNA	2-5μl
T4 ligase	1μl
Pure water	5-8μl
Total	**10μl**

I ligated at room temperature for only half an hour, due to the high efficiency of the NEB ligase, and then immediately transformed competent DH5α *E. Coli* with the ligated vector according to the following protocol, which I adopted for all transformations I performed:

1. Thaw competent cells on ice.

2. Pipette the 10μl ligation reaction into 50-75μl competent cells.

3. Sit on ice for 20 minutes.

cleave the fusion protein from the desired immunogen.

[10]PCR extension times were doubled when using Pfu as it is a slower enzyme than *Taq* due to its 3' exonuclease proofreading activity.

4. Heat-shock at 42°C for 90 seconds.

5. Place on ice for 2 minutes.

6. Plate immediately onto appropriate agar growth medium under sterile conditions.

With pGEM-T Easy I used standard LB-agar-ampicillin with added IPTG and X-Gal for blue-white screening of colonies [Sambrook et al., 1989]. I grew the bacteria at 37°C over-night and put them in the fridge the next morning. In the afternoon I picked a well isolated white colony to seed 3-5ml of selective growth medium (LB-ampicillin for pGEM-T Easy) under sterile conditions. I repeated this for several colonies from each successful cloning plate. I grew the seeded media over-night at 37°C. The next morning I took a small amount of medium separately from each successful growth to make a glycerol stock (25% glycerol final) for storage of successful clones at -80°C. I used the remaining medium from each successful growth to extract the plasmid DNA using a Miniprep (or Maxiprep if more DNA was required) kit (Qiagen). I eluted the plasmid in 50μl pure water to most often yield 50ng/μl DNA. I digested 2μl of this plasmid DNA with appropriate restriction enzyme(s) (*EcoRI* for pGEM-T Easy) and ran a gel to visualize the cut plasmid and insert. I then bi-directionally sequenced plasmid DNA containing appropriately sized insert.

I sub-cloned inserts with appropriate sequence into the expression vectors. I prepared the vectors by performing a sequential digest first with *SalI* and then with *EcoRI*[11]. Both digests were performed in the supplied *SalI* buffer to avoid loss of DNA during an intermediate purification step. It is hard to cleave DNA ends with little overhang using *SalI* so it was used first while the plasmid was still circular. The insert was cut out of pGEM-T Easy using a double-digest as there was plenty of overhang at each restriction site. I then cloned the inserts separately into each expression vector as outlined above using the appropriate bacterial strain, media and antibiotic selection.

[11] *EcoRI* is a very efficient enzyme that can easily cut DNA at a restriction site with little or no overhang and also works under a number of different salt conditions. See the New England Biolabs catalogue for a useful reference on restriction digests.

2.16 Prokaryotic protein expression

I expressed NPIP-GST and -MBP fusion proteins by transforming the appropriate bacterial strain. For expression in pGEX I used BL21 *E. Coli.* For pMAL and pEGFP I used DH5α *E. Coli.* I picked a single colony from a plate of transformed cells containing the expression plasmid with sequenced insert after sub-cloning. I grew the colony in 50ml growth medium, with antibiotic selection, at 37°C over-night and added the growth to a further 500ml of the same medium the next morning. I performed a restriction digest to check the size of the insert. The medium was grown to a density of 0.5-1.0 by measuring the absorbance at 600nm (A_{600}) using a spectrophotometer (fresh medium was used to zero the meter). The *E.Coli* were then induced to produce protein with 0.2mM IPTG and grown at 30°C, to minimise degradation, for four hours. I spun down the cells in 2 250ml containers at 6,000 rpm for 20 minutes in a centrifuge, discarded the medium and froze the pellets at -80°C over-night. The next day I thawed the pellets and re-suspended the cells on ice in a total of 15ml PBS with 1mM AEBSF (Calbiochem), a stable non-toxic protease inhibitor. I sonicated the cells to complete the lysis (that started by freezing them) with six 10 second bursts of ultra-high frequency sound with an amplitude of 14-15μm. I used the protein for Western blotting (see chapter 5) to characterize the antisera and subsequently to compare the expression of native NPIP in human retinal extract.

I expressed GFP-NPIP fusion proteins by transient transfection of human neuroblastoma (SK-N-SH) cells using the Lipofectamine™ kit from Invitrogen. I followed the instructions from the kit and used 100ng of plasmid DNA per well of an 8-chamber microscope slide to transfect cells cultured at $\approx$25% confluence. I also transfected cells with empty pEGFP-C2 plasmid as a positive control and cultured un-transfected cells on the same slide as a negative control. All experiments were carried out in duplicate. I checked the cells for confluence and expression of fluorescent fusion protein after over-night growth. After 24 hours I fixed the cells in 3.7% paraformaldehyde and imaged them by confocal microscopy.

Chapter 3

Microarray analysis

3.1 Introduction

In the 1999 *Nature Genetics* Supplement, "The Chipping Forecast", Southern et al. begin:

> *The DNA microarray is the latest in a line of techniques to exploit a potent feature of the DNA duplex—the sequence complementarity of the two strands. It is remarkable that a molecule of such great structural complexity can reassemble with perfect fidelity from the separated strands.*

The introduction of solid supports has greatly increased the range of applications for nucleic acid hybridization.

The starting point, for the use of duplex formation as a tool in molecular biology, was the observation that single-stranded DNA binds strongly to a nitrocellulose membrane in a way that prevents the strands from re-associating with each other, but does permit hybridization to complementary RNA [Gillespie and Spiegelman, 1965]. This was the direct precursor of the 'blotting' methods, the first of which, Southern blotting [1975], combined gel separation of restriction digests with hybridization on filters.

Historically, most applications of blotting have combined a single labelled, purified oligonucleotide in the liquid phase with a complex mixture of polynucleotides attached to a solid support [Duggan et al., 1999]. For example, transcript abundance is often assessed

by immobilizing mRNA or total RNA (electrophoretically separated as in 'Northern blotting') on membranes and then incubating with a radioactively labelled, gene-specific target (see 2.3.2). If multiple RNA samples are immobilized on the same membrane, information about the quantity of a particular message in each sample can be obtained.

The method of 'dot-blotting' [Kafatos et al., 1979] was developed for rapidly determining the relative concentrations of nucleic acids in a mixture. Subsequent automation and miniaturization of the dot-blot showed how hybridization could be used on a large scale to exploit the data emerging from genome programmes [Lennon and Lehrach, 1991].

The main advance from dot-blots to microarrays is the use of an impermeable rigid support, such as glass, which has practical advantages over porous membranes [Khrapko et al., 1989]. As liquid cannot penetrate the surface of the support, target nucleic acids have immediate access to the probes without diffusing into pores. This enhances the rate of hybridization. The washing step that follows is also unimpeded by diffusion, speeding up the procedure and improving reproducibility. The flatness, rigidity and transparency of glass supports improve image acquisition and processing, as the locations of the probes are much better defined than on a flexible membrane; high resolution is critical for the small feature sizes that can be achieved on microarrays. These practical benefits apply to arrays of cDNAs, PCR products and synthetic oligonucleotides.

The two main platforms for high-density microarrays are:

1. Direct synthesis of oligonucleotides, usually 15–25-mers, on glass substrates using photolithography [Fodor et al., 1991].

2. Deposition of DNA samples on to glass substrates using a spotting robot [Schena et al., 1995].

One advantage of spotted microarrays over direct synthesis microarrays is greater versatility. Researchers can make and design their own microarrays, selecting the DNA sources to spot. These arrays are also lower cost, especially for studies involving a large number of experiments. Amplified products of cDNA inserts from expressed sequence tags (ESTs) are commonly used as a source of samples for spotted microarrays [Schena

et al., 1995], although several technical difficulties are often associated with using this source [Finkelstein et al., 2002]. Firstly, mismatches of cDNA clones and EST sequences in the database can be as high as 38% [Knight, 2001] due to tracking errors or contamination. Secondly, the tracking and quality tests of amplified products can be tedious and difficult to manage. Thirdly, cDNAs may not distinguish among members of multigene families, related genes, and alternatively spliced genes. Finally, current EST collections represent only 25–50% of the predicted genes in a genome.

An alternative source of DNA samples for spotted microarrays are long synthetic oligos (50-70mers) designed from ESTs or annotated genome sequences. Synthetic oligos have advantages over cDNAs because they avoid the need to track clones and amplicons, and they can be designed to have a uniform sequence length (ensuring more uniform hybridization) and high specificity for distinguishing related DNA sequences. Sets of long oligos have been developed using information from all annotated genes of complete genome sequences for several model organisms, including yeast, Drosophila , human, mouse, and Arabidopsis `www.operon.com`.

Spotted arrays are generally used with two coloured fluorescent probes, whereas arrays produced by direct synthesis, for example GeneChips (Affymetrix), use only a single intensity reading meaning that only one probe rather than two can be hybridized in a single experiment. In two-colour microarray expression analysis many gene-specific polynucleotides derived from the 3′ end of RNA transcripts are individually arrayed on a single matrix. This matrix is then *simultaneously* hybridized with fluorescently labelled cDNA representations of RNA from two separate pools. This allows the relative amount of transcript present in each pool to be determined by measuring fluorescence at two wavelengths.

For example, if we have an array with $array$ distinct elements and compare fluorescence values (in arbitrary units) for query and reference samples, $Cy3$ and $Cy5$ respectively (for the green and red fluorescent dyes commonly used in two-colour array experiments), then for the ith gene (where i is an index running over all the arrayed genes

from 1 to $array$), the expression ratio M can be written as

$$M_i = \frac{Cy3_i}{Cy5_i}. \tag{3.1}$$

Although *ratios* provide an intuitive measure of expression changes, they have the disadvantage of treating over- and under-expressed genes differently. Genes over-expressed by a factor of 2 have an expression ratio of 2, whereas those under-expressed by the same factor have an expression ratio of 0.5.

The most widely used alternative to the simple ratio is the logarithm base 2 transformation of the ratio, which has the advantage of producing a continuous spectrum of values, treating over- and under-expressed genes in a similar fashion. Logarithms treat numbers and their reciprocals symmetrically: $log_2(1) = 0$, $log_2(2) = 1$, $log_2(1/2) = -1$, $log_2(4) = 2$, $log_2(1/4) = -2$ and so on. The logarithms of the expression ratios are also treated symmetrically, so that a gene over-expressed by a factor of 2 has a log_2 ratio of 1, a gene under-expressed by a factor of 2 has a log_2 ratio of -1, and a gene expressed at a constant level (with a ratio of 1) has a log_2 ratio equal to zero.

Differential gene expression data from microarray experiments are usually calibrated (normalized) to account for variations in sample treatment, RNA quantification, labelling and dye efficiency, detection, and other systematic biases in the measured expression levels (reviewed in Quackenbush [2002]). Typically, normalization adjusts the individual fluorescence intensities to balance them appropriately so that meaningful biological comparisons can be made. Conceptually, normalization is similar to adjusting expression levels measured by Northern blotting or quantitative PCR (see chapter 4) relative to the expression of one or more reference genes, whose levels are assumed to be constant between samples.

There are many approaches to normalizing expression levels. Some, such as *total intensity normalization*, are based on simple assumptions. Given equal *amounts* (weights) of labelled cDNA in each sample, one assumption is that the *number* of labelled cDNA molecules in each sample is also the same. A second assumption is that the arrayed

elements represent a random sampling of the genes in the organism. If the arrayed genes are selected to represent only those we know will change, then we will likely over- or under-sample the genes in one of the biological samples being compared. However, if the array contains a large enough assortment of random genes, we do not expect to see such a bias. Consequently, approximately the same number of labelled molecules from each sample should hybridize to the arrays and, therefore, the total hybridization intensities summed over all elements in the array should be the same for each sample.

Using the above assumptions[1] a *normalization factor* N is calculated by dividing the summed intensities for each channel:

$$N = \frac{\sum_{i=1}^{array} Cy3}{\sum_{i=1}^{array} Cy5}. \tag{3.2}$$

One or both intensities can then be appropriately scaled, for example

$$Cy5'_i = N \times Cy5_i \text{ and } Cy3'_i = Cy3_i \tag{3.3}$$

so that the normalized expression ratio for each element becomes

$$M'_i = \frac{Cy3'_i}{Cy5'_i} = \frac{1}{N}\frac{Cy3_i}{Cy5_i}, \tag{3.4}$$

which adjusts each ratio such that the mean ratio is equal to 1. This is equivalent to subtracting the *normalization factor* from the logarithm of the expression ratio,

$$M'_i = log_2(M_i) - log_2(N) \tag{3.5}$$

which results in a mean log_2 ratio of zero.

There are many variations on this type of normalization, including scaling the individual intensities so that the mean or median intensity is the same within a single array

[1] Different amounts of cDNA in each sample *can* be taken into account by total intensity normalization. Equal amounts are assumed for explanatory purposes.

or across all arrays, or using a selected subset of genes to normalize against. In addition to total intensity normalization there are a number of alternative approaches to normalizing expression ratios, including linear regression analysis, log centering, rank invariant methods [Tseng et al., 2001] and ratio statistics [Chen et al., 1997]. However, none of these approaches takes into account systematic bias that appears in the data [Yang et al., 2002a,b], most commonly as a deviation from zero for low-intensity spots.

The *variability* in the log_2 ratios also increases as the hybridization intensity decreases: relative error increases at lower intensities where the signal approaches background. Commonly used approaches to address this problem are to use only array elements with intensities that are statistically significantly different from background, and to use absolute lower thresholds for 'acceptable' array elements or percentage based cut-offs. Conversely, at the high end of the intensity spectrum – where the array elements saturate the fluorescence detector – saturated pixels can be eliminated in the image-processing step or a maximum 'acceptable' value can be set.

However, there are a number of normalization techniques that take into account systematic bias without pre-processing the intensity data. One method proposed by Yang et al. [2002b] to remove intensity-dependent effects in the log_2 ratios is locally weighted linear regression analysis (*lowess*).

Another approach to deal with the problem of systematic intensity bias is to adjust the log_2 ratios so that the amount of statistical *variance* is equal for spots of all intensities [Yang et al., 2002b; Huber et al., 2002]. This has the advantage of being able to adjust both for differences between regions of an array or between arrays.

Lowess, and other methods of normalizing intensity, can be applied either globally or locally. For spotted arrays, local normalization is often applied to each group of array elements deposited by a single pen, as in the *print-tip normalization* methods used in Terry Speed's lab [Yang et al., 2002b; Smyth and Speed, 2003]. Local normalization can help correct for systematic *spatial* variation in the array, including inconsistencies among pens used to make the array, variability in the slide surface, and slight local differences in

hybridization conditions. However, this is not important in all types of array.

Microarray experiments have previously been used in the retina to identify both disease-causing (autosomal dominant retinitis pigmentosa [Kennan et al., 2002]) and major regulatory (homeobox Crx [Livesey et al., 2000]) genes. However, many microarray experiments suffer from too few replicates and lack of statistical analysis.

3.2 Objectives

I planned to make use of microarray technology to look for differences in gene expression between central and peripheral retina in a large number of genes. My aim was to perform multiple microarray experiments using the human macular resource (chapter 1) I procured. An important objective was to analyse the data obtained from microarray experiments in a way that discriminated between genes well enough to choose a handful for further study. Analysing the data from first principles, rather than relying on proprietary software, was part of this objective. I planned to compare different methods of analysis and choose those most appropriate to my particular experiments.

3.3 Results

3.3.1 Resources

I found over 50 retina-related genes amongst the 13,899 unique probes on the Array Centre microarrays (see Appendix B Table B.1). I used six of these slides in separate hybridization experiments to compare gene expression in the central and peripheral retina. Messenger RNA isolated from 24 separate donor eyes (see section 2.3) was used for microarray and subsequent quantitative PCR analysis. The donor age varied from forty to seventy-seven years and the sex distribution was approximately even. No signs of retinal disease were observed in any of the eyes used, nor was any ocular disease mentioned on the consent forms. All eyes were placed at four degrees Celsius post-enucleation and

many were additionally bathed in RNAlater (Ambion, Texas) at this time. Most donors had been cooled in the morgue prior to enucleation. The time between death and either dissection or RNAlater treatment ranged from four to twelve hours.

Two separate RNA samples from the mid-peripheral retinas of separate donors were chosen at random from the best quality RNA samples obtained using RNAlater and the column extraction method. The peripheral retina 1 sample originated from a 72 year old woman who died from a cardiac arrest secondary to lung cancer (lane 2, autoreffig:peri) and whose eyes were enucleated and bathed in RNAlater within 7 hours of death. The peripheral retina 2 sample originated from a 57 year old man who died from a stroke secondary to carotid artery stenosis (lane 2, autoreffig:peri) and whose eyes were enucleated and bathed in RNAlater within 6 hours of death. The two pooled 4mm punches and ten pooled 2mm punches (Figure 1.17, left and right panels respectively) used were extracted using the modified TRIZOL method. Dye-swap experiments were not performed as the amount of foveal RNA was a limiting factor.

3.3.2 Experiments

1. peripheral retina 1 *vs.* peripheral retina 1
2. peripheral retina 1 *vs.* peripheral retina 2
3. 4mm pooled macula *vs.* peripheral retina 1
4. 4mm pooled macula *vs.* peripheral retina 2
5. 2mm pooled foveo-macula *vs.* peripheral retina 1
6. 2mm pooled foveo-macula *vs.* peripheral retina 2

Each experiment was carried out on a new microarray slide that had been subject to quality control. The scanned slides consisted of two channels of data each – one for Cy3 and one for Cy5 (see Figure 3.1). High resolution TIFF files (≈25 Megabytes per channel) were used to align grids corresponding with the spots on the arrays and numerical spot

fluorescence values were successfully extracted using ImaGene. The custom R scripts then successfully parsed the data and the map of genes on the array into text files.

3.3.3 Optimisation

Diagnostic M *vs.* A (see Equation 3.6) plots using data from each of these 6 experiments displayed varying amounts of bias with the different normalization techniques used. This analysis was particularly useful for experiment 1 – a self-self control in which two aliquots of the same RNA sample were labelled with different dyes and hybridized competitively on the array (see /autoreffig:reanalysis). Non-normalized log-ratios produced the most bias. *Intensity normalization* (using the median expression as the normalization factor) resulted in a simple shift of the log data in the y-axis towards the origin, but did not otherwise change the distribution of the data. The *lowess* and *print-tip normalization* graphs were very similar to the intensity normalized graphs. A substantial reduction in bias was seen with the SMA scaling algorithm and variance stabilisation, with the funnel effect of low-expressed genes having exaggerated expression ratios most markedly reduced with the latter method.

The top twenty most highly differentially expressed genes were calculated using each normalization method (see Table 3.1) for experiment five. The list was identical using either no normalization (simple log ratios) or the median expression as a normalization factor. The other lists were similar, with some positional differences in the top twenty and a few omissions and additions. The SMA scale algorithm produced a top twenty list quite dissimilar to any other.

The expression ratios for rod and red/green cone opsin were mined from the data obtained for experiments 5 and 6 (2mm foveo-macula *vs.* peripheral retina) using each of the normalization methods and means calculated for the four spots (two on each slide) to verify the direction and scale of opsin expression using the different analysis methods. All the analysis methods reported red/green cone opsin over-expression in foveo-macula 2mm and rhodopsin over-expression in peripheral retina except the non-normalized log

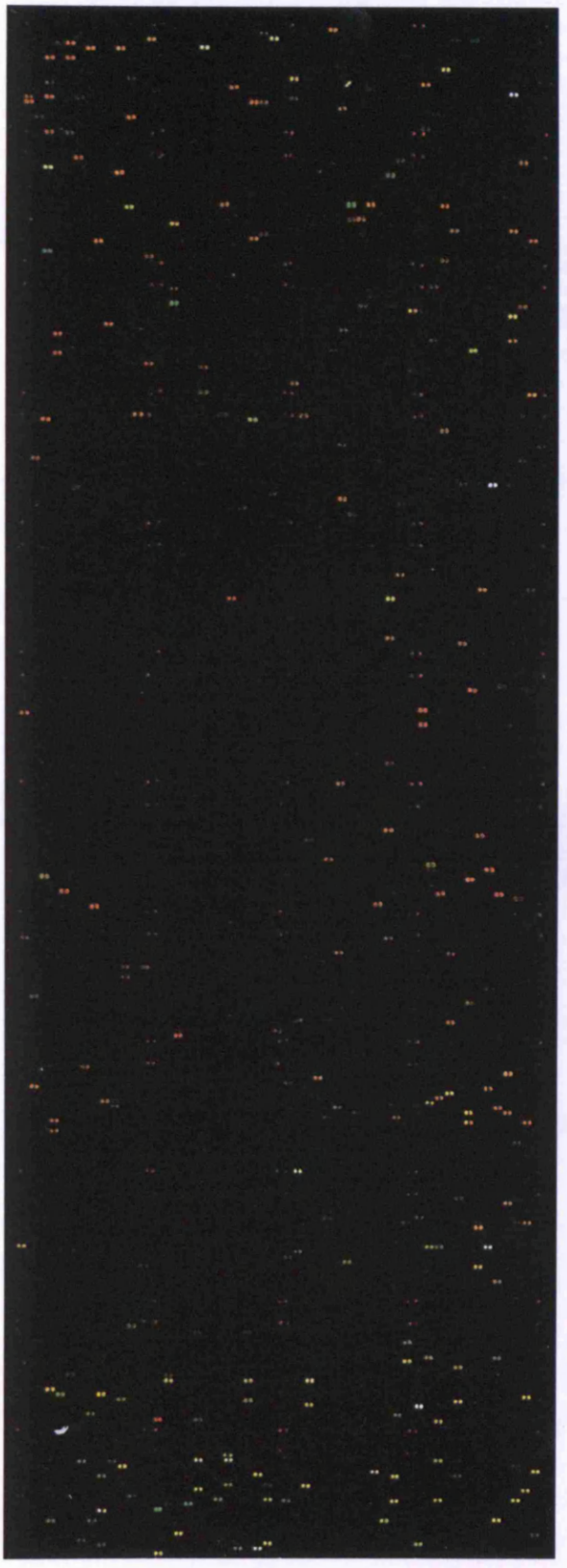


Figure 3.1: Low resolution merged two-channel image of microarray slide 6 – 2mm pooled foveo-macula (Cy3, green) *vs*. peripheral retina 2 (Cy5, red).

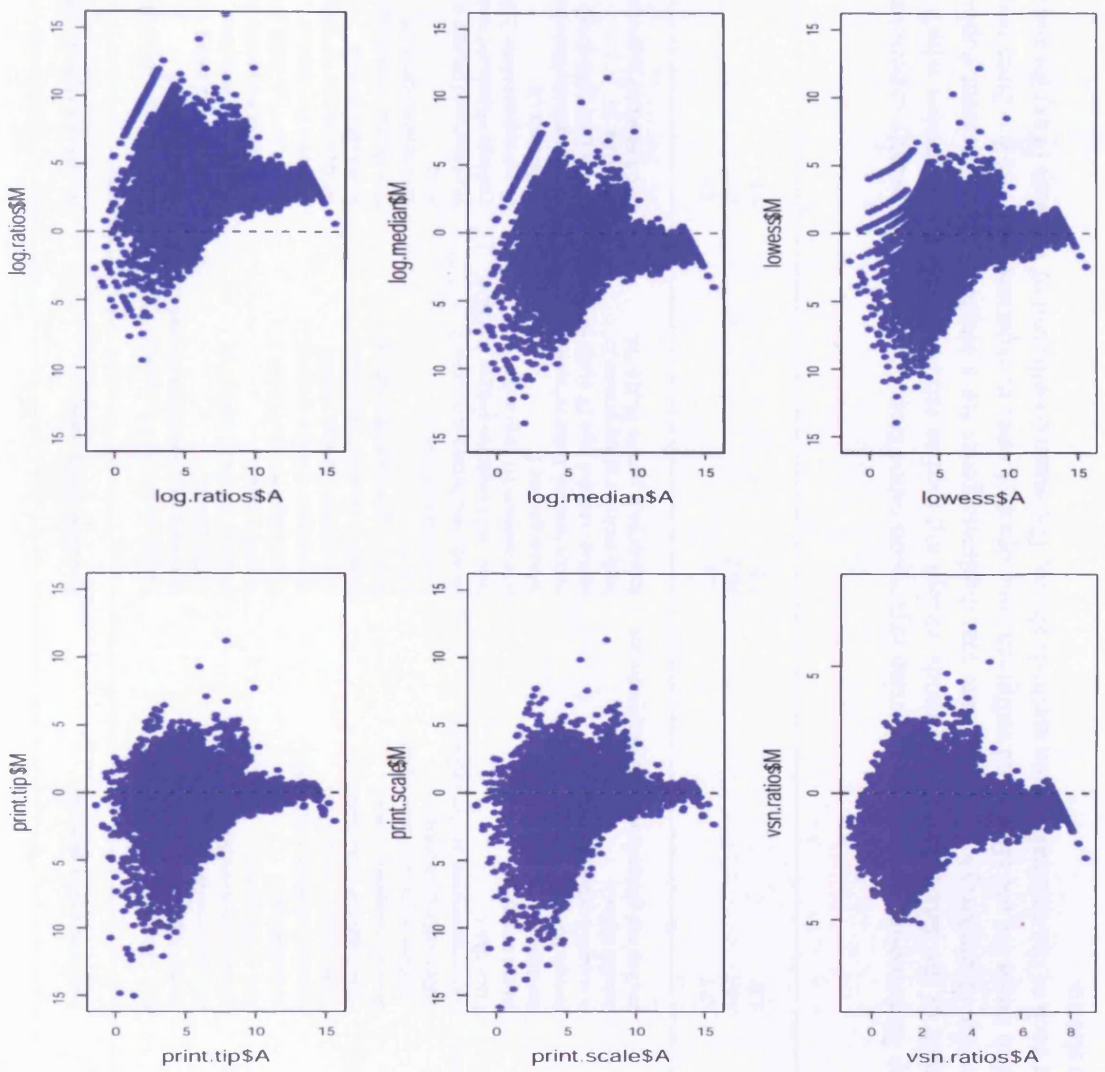


Figure 3.2: Diagnostic microarray *M vs.* A plots using data from the self-self (peripheral retina 1 *vs.* peripheral retina 1) duplicate microarray experiment.

	Log ratio	Log ratio (SMA lowess)	Log ratio (SMA print tip)	Log ratio (SMA scale)	Variance stabilised ratio
	neurofilament 3	neurofilament 3	FAT tumor suppressor homolog 2	Human clone C3 CHL1 protein	C3H-type zinc finger protein
	methyl-CpG binding domain 4	ubiquitin thiolesterase	methyl-CpG binding domain 4	amiloride-sensitive cation channel 3 testis	neurofilament 3
	FAT tumor suppressor homolog 2	ubiquitin thiolesterase	ubiquitin thiolesterase	NESH protein	methyl-CpG binding domain 4
	ubiquitin thiolesterase	FAT tumor suppressor homolog 2	ubiquitin thiolesterase	papillary renal cell carcinoma	ubiquitin thiolesterase
	ubiquitin thiolesterase	methyl-CpG binding domain 4	KIAA1001 protein	crystallin beta A1	FAT tumor suppressor homolog 2
	KIAA1001 protein	KIAA1001 protein	neurofilament 3	hypothetical protein F23149 1	C3H-type zinc finger protein
	hypothetical protein FLJ11029	hypothetical protein FLJ11029	hypothetical protein FLJ11029	tumor necrosis factor receptor 12	ubiquitin thiolesterase
	C3H-type zinc finger protein	Homo sapiens clone 23860	Homo sapiens clone 23860	ribonuclease HI large subunit	far-diphosphate far-transferase 1
	Homo sapiens clone 23860	cartilage oligomeric matrix protein	Homo sapiens clone 24894	papillary renal cell carcinoma	KIAA1001 protein
	far-diphosphate far-transferase 1	dev diff enhancing factor 2	dev diff enhancing factor 2	beta-1 3-glucuronyltransferase 1	hypothetical protein FLJ11029
	cartilage oligomeric matrix protein	Dmx-like 1	cartilage oligomeric matrix protein	GS15	dopamine receptor D5
	dev diff enhancing factor 2	Homo sapiens clone 24894	cDNA DKFZp564D042	EUROIMAGE 35907	NF-L
	Dmx-like 1	Rab gergertransferase beta subunit	Rab gergertransferase beta subunit	parathyroid hormone receptor 2	yes-associated protein 65 kDa
	Homo sapiens clone 24894	far-diphosphate far-transferase 1	Dmx-like 1	renal cell carcinoma antigen RAGE-4	cartilage oligomeric matrix protein
	Rab gergertransferase beta subunit	cDNA DKFZp564D042	paired box gene 6	ribonuclease HI large subunit	Homo sapiens clone 23860
	C3H-type zinc finger protein	paired box gene 6	far-diphosphate far-transferase 1	tubulin gamma 1	homeo box C6
	cDNA DKFZp564D042	homeo box C6	homeo box C6	tumor necrosis factor receptor 12	Rab gergertransferase beta subunit
	paired box gene 6	mannose-6-phosphate receptor	mannose-6-phosphate receptor	sodium channel type IX alpha polypeptide	cDNA DKFZp564D042
	mannose-6-phosphate receptor	zinc finger protein 237	tubulin alpha 2	high-mobility group protein 2	Dmx-like 1
	homeo box C6	tubulin alpha 2	proline and glutamic acid rich nuclear protein	hypothetical protein FLJ10734	PEG8/IGF2AS imprinting gene
Rod	-4.2	-3.3	-2.7	-1.5	-0.9
Cone	-1.4	+0.9	+0.7	+0.7	+1.2
Var	1.4	0.8	1.0	1.2	0.5

Table 3.1: Diagnostic microarray data. **Upper section**: Top 20 foveo-macula:peripheral retina expression ratios from (duplicate) microarray experiment 5 – 2mm pooled foveo-macula *vs.* peripheral retina 1 – for each of the data calibration methods, except log median ratio intensity normalization which has the same top 20 as the log ratio method. **SMA** = Statistics for MicroArrays package. Note that duplicate genes for a single method represent duplicate spots within the microarray. **Lower section**: Mean expression ratios for rod (**Rod**) and red/green cone opsin (**Cone**) in experiments 5 and 6 – 2mm pooled foveo-macula *vs.* peripheral retina 1 and 2 respectively – for each of the data calibration methods above. The mean co-efficient of variance (**Var**) for rod and cone opsin using the different methods of calibration is also shown.

ratio, which also reported red/green cone opsin over-expression in the peripheral retina. This result was also seen in an initial analysis of the data using GeneSpring at the time. Rhodopsin over-expression in peripheral retina was reported to be substantially more than red/green cone opsin over-expression in 2mm fovea using all but the variance stabilisation algorithm. Variance stabilisation reported red/green cone opsin over-expression in 2mm fovea to be a third greater than rhodopsin over-expression in the peripheral retina.

3.3.4 Data analysis

All the processed microarray data were subsequently re-analysed using both non-normalized log-ratios (as a comparison) and *variance stabilization*. I applied these two transformations separately to each slide and plotted the data as the expression ratio between channels M *vs.* the rank of the mean expression A defined by

$$\begin{aligned} M &= \frac{h(Cy3_i)}{h(Cy5_i)} \\ A &= \frac{h(Cy3_i) + h(Cy5_i)}{2} \end{aligned} \tag{3.6}$$

where $h(x)$ is defined by equation 3.7 or 3.10. The ranks rather than the absolute values of mean expression were plotted as it makes it easier to visualise differential expression (by reducing the central bulge in the data due to most genes being of mid-level expression).

Figure 3.3 shows M *vs.* rank A plots for duplicate data from experiments 1-6. In the upper graph of each panel, the function:

$$h(x) = log_2(x) \tag{3.7}$$

was applied to the data. Thus, M was the log (base two) expression ratio of Cy3:Cy5 for each non-control spot on the array:

$$M_i = \log_2(Cy5_i) - \log_2(Cy3_i) \equiv \log_2(Cy5_i/Cy3_i) \tag{3.8}$$

Therefore, +1 represents two-fold over and -1 two-fold under-expression.

Similarly

$$A_i = \frac{\log_2(Cy3_i) + \log_2(Cy5_i)}{2} \equiv log_2\sqrt{Cy3_i \times Cy5_i} \tag{3.9}$$

which is the arithmetic mean of the log expression (or the log geometric mean of expression).

In the lower graph of each panel, $h(x)$ is the *variance stabilized* data [Huber et al., 2002]. The values for the expression ratio M calculated using the *variance stabilization* function are equivalent to log-ratios for high intensity microarray spots. However, *variance stabilization* is designed to reduce the noise in genes with lower mean expression A – a small change in intensity results in a relatively large change in differential expression for these genes when log-ratios are used. *Variance stabilization* is achieved by applying the function

$$h(x) = \sinh^{-1}(a + bx) \tag{3.10}$$

where a and b are constants calculated from the expression data according to the variance stabilisation algorithm [Huber et al., 2002].

The difference in the log-transformed (upper graph) and variance stabilised (lower graph) data in figure 3.3 can be clearly seen. The log-transformation resulted in many low-expressed genes with large differences in expression. These funnel-shaped plots are typical for logarithmic transformations of microarray data [Chen et al., 1997].

The variance stabilized data produced much flatter graphs that were more centred around the origin, especially for the control experiments (1&2). A characteristic central bulge of differentially expressed genes at mid-level expression appears in the central *vs.* peripheral experiments, except for experiment 3. There are also a number of highly expressed genes which also appear highly *differentially* expressed in experiments 4-6.

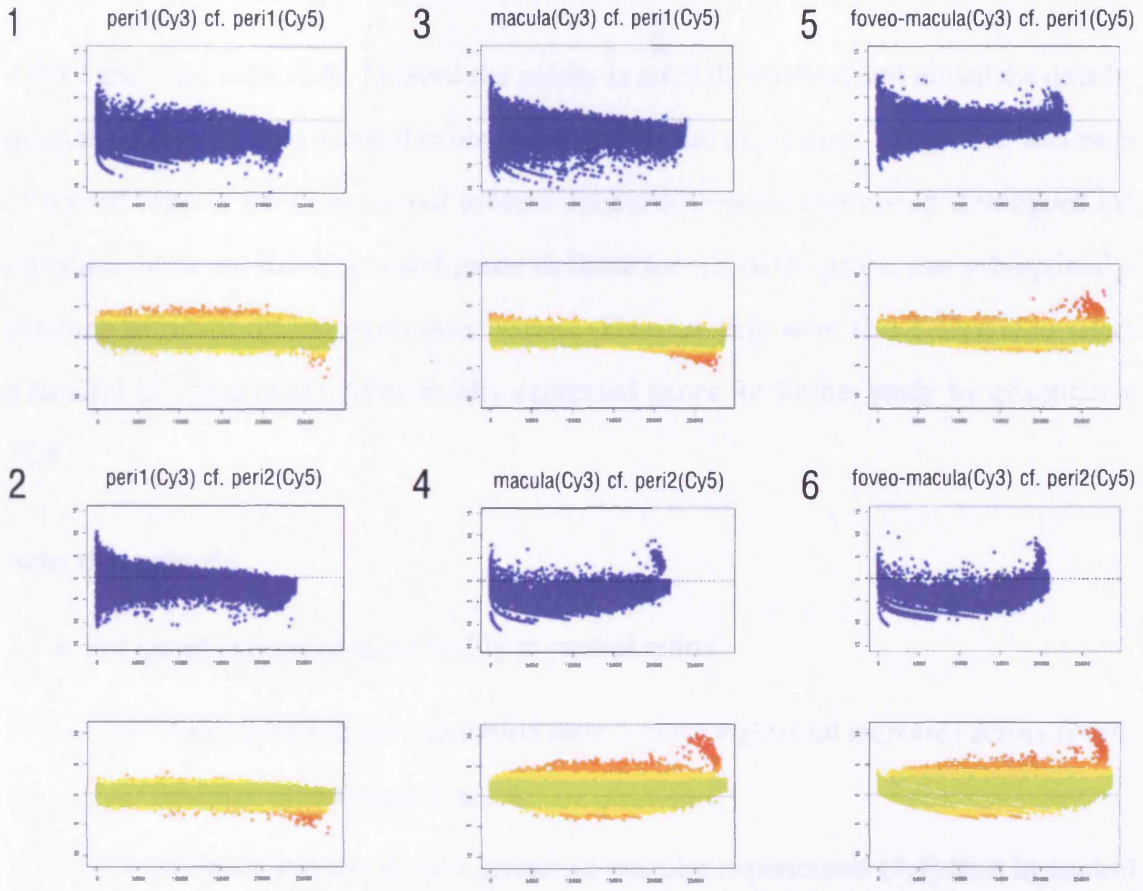


Figure 3.3: Relative *vs.* rank mean expression ratios in six microarray experiments. The upper graphs show log-ratios whilst the lower graphs show variance stabilised ratios.

3.3.5 Data mining

After I analysed each slide, I placed the results in an SQL database and mined the data by querying for genes that showed evidence of differential expression. Genes that met each of several criteria were considered to show robust differential expression. I designed the criteria to filter out low-expressed genes or those for which the probe was sub-optimal – resulting in relatively low expression values. These criteria were also designed to select a handful of consistently differentially expressed genes for further study by quantitative PCR.

Selection criteria

- For genes expressed more highly in central retina:
  - Variance stabilised expression ratio > 3 ($\approx$ eight-fold increase) across foveo-macular experiments 5 and 6.
  - Variance stabilised ratio greater in macular experiments (3,4) than in control experiments (1,2).
  - Variance stabilised ratio greater in foveo-macular experiments (5,6) than in macular experiments (3,4).
- For genes more highly expressed in peripheral retina:
  - Variance stabilised expression ratio ≤ -2 ($\approx$ four-fold decrease) across experiments 5 and 6.
  - Variation in ratios of < 1 across control experiments 1 and 2.
- For all genes:
  - Absolute fluorescence of probes > 600 above background in at least 2 channels.

Table 3.2 shows background-corrected fluorescence and variance stabilised data for two control genes, rod and red/green cone opsin. The red/green cone opsin probe was

designed to long wave (red) cone opsin. However, red and green opsin are too similar to be distinguished by hybridization (or even RT-PCR).

Table 3.3 shows fluorescence and variance stabilised data for two genes, S-antigen and nuclear pore complex interacting protein, that met all of my criteria for differentially expressed genes.

Table 3.4 shows a summary of data from probes that met all of the above criteria, as well as from the rod and long wave cone opsin probes.

3.4 Discussion

The microarray slides were designed from a set of 13,899 cDNAs, each corresponding to a unique human UniGene cluster. Of these, over 50 genes were related to the retina (see Appendix B Table B.1). This allowed me to look at the expression of a large number of genes of potential importance to both precise vision and macular degeneration.

Intensity normalization, using the median expression as the normalization factor, did centre the expression data around the origin, although exactly the same intensity biases were still evident as in the non-normalized data (see Figure 3.2. These biases include the non-linear trends seen especially in experiments 4 and 6 (see 3.3) and most notably the large amount of variation in ratios for genes with low measured hybridization intensity.

Intensity normalization was as far as most commercially available software went towards normalization of array data. In fact, it is possible that the version of GeneSpring used at the time did not employ any normalization at all, given that both GeneSpring and the non-normalized log-ratios both reported *under*-expression of red/green cone opsin in the foveo-macula. The normalization algorithm used in GeneSpring, the only commercially available software accessible to me at the time, was not publicly available. This was a factor in deciding to analyse the microarray data using non-propietary but well-researched algorithms.

Print-tip normalization, using the SMA package, did not appear to improve the overall shape and hence bias of the data significantly. Perhaps the high degree of quality control

Rhodopsin	Cy3	Cy5	Vsn3	Vsn5	Vsnr	ΔVsnr
Self-self (peri2B)	661	11504	5.75	5.63	0.12	
	410	8737	5.27	5.35	-0.08	
Intra	536	10120	5.51	5.49	0.02	0.10
Peri1-peri2B	337	1844	3.19	2.41	-0.78	
	175	1787	3.16	1.73	-1.43	
Inter	256	1816	3.17	2.07	-1.10	1.10
Control	396	5968	4.34	3.78	-0.54	0.60
Fovea4mm-peri2B	6	385	1.21	2.06	-0.84	
	-3	297	1.00	1.91	-0.91	
Fovea4mm-periL1	120	916	4.68	3.81	0.87	
	88	780	4.36	3.64	0.73	
Fovea4mm	53	594	2.82	2.85	-0.04	0.84
Fovea2mm-periL1	372	2807	3.00	4.21	-1.22	
	456	2989	3.20	4.28	-1.07	
Fovea2mm-peri2B	12	603	2.94	3.74	-0.80	
	15	635	3.18	3.79	-0.61	
Fovea2mm	213	1758	3.08	4.01	-0.92	0.22

R/g cone opsin	Cy3	Cy5	Vsn3	Vsn5	Vsnr	ΔVsnr
Self-self (peri2B)	26	218	2.60	1.77	0.84	
	4	266	1.21	1.94	-0.73	
Intra	15	242	1.91	1.85	0.05	0.78
Peri1-peri2B	121	120	0.64	1.36	0.71	
	38	407	1.68	0.39	-1.29	
Inter	80	264	1.16	0.88	-0.29	1.00
Control	47	253	1.54	1.37	-0.12	0.89
Fovea4mm-peri2B	65	432	2.15	2.13	0.02	
	32	699	1.73	2.46	-0.73	
Fovea4mm-periL1	4	88	1.15	0.61	0.54	
	3	58	0.86	-0.15	1.01	
Fovea4mm	26	320	1.47	1.26	0.21	0.57
Fovea2mm-periL1	374	194	3.01	1.42	1.58	
	316	79	2.84	0.53	2.31	
Fovea2mm-peri2B	11	165	2.89	2.16	0.73	
	5	130	2.03	1.80	0.23	
Fovea2mm	177	142	2.69	1.48	1.21	0.73

Table 3.2: Background-corrected fluorescence (*Cy3* and *Cy5*) and variance stabilized (*Vsn3* and *Vsn5*) data for rod and red/green cone opsin. *Vsnr* is the ratio of *Vsn3:Vsn5*. The samples used for each experiment are listed in the leftmost column. Data for duplicate probe spots are shown in consecutive rows. Means for each group of experiments are shown in the rows entitled *Intra* (for the self-self control), *Inter* (for the peri1-peri2B control), *Control* (for the combined controls), *Fovea4mm* (for both 4mm fovea experiments) and *Fovea2mm*. $\Delta Vsnr$ is the average deviation of Vsnr from zero in the control experiments and from the mean Vsnr in the non-control experiments.

S-antigen	Cy3	Cy5	Vsn3	Vsn5	Vsnr	ΔVsnr
Self-self (peri2B)	557	8678	5.57	5.34	0.23	
	494	7321	5.45	5.18	0.28	
Intra	525	7999	5.51	5.26	0.25	0.25
Peri1-peri2B	1323	4967	4.19	3.80	-0.38	
	2056	8439	4.72	4.24	-0.47	
Inter	1690	6703	4.45	4.02	-0.43	0.43
Control	1107	7351	4.98	4.64	-0.09	0.34
Fovea4mm-peri2B	65	494	2.15	2.22	-0.07	
	30	90	1.69	1.43	0.26	
Fovea4mm-periL1	12	340	2.28	2.69	-0.41	
	38	204	3.50	2.02	1.48	
Fovea4mm	36	282	2.41	2.09	0.32	0.58
Fovea2mm-periL1	142	769	2.04	2.89	-0.85	
	87	904	1.57	3.06	-1.49	
Fovea2mm-peri2B	0	207	-0.61	2.47	-3.08	
	1	268	0.23	2.80	-2.57	
Fovea2mm	57	537	0.81	2.80	-2.00	0.83

NPIP	Cy3	Cy5	Vsn3	Vsn5	Vsnr	ΔVsnr
Self-self (peri2B)	116	1810	4.02	3.79	0.24	
	183	2256	4.47	4.00	0.47	
Intra	150	2033	4.25	3.89	0.35	0.35
Peri1-peri2B	871	1972	3.26	3.38	0.12	
	768	1436	2.94	3.25	0.31	
Inter	819	1704	3.10	3.31	0.21	0.21
Control	484	1868	3.67	3.60	0.28	0.28
Fovea4mm-peri2B	397	4167	3.65	4.02	-0.37	
	563	5494	3.98	4.28	-0.30	
Fovea4mm-periL1	968	1584	6.78	4.39	2.39	
	1072	1400	6.88	4.26	2.62	
Fovea4mm	750	3161	5.32	4.24	1.08	1.42
Fovea2mm-periL1	4004	1675	5.38	3.69	1.69	
	3700	1392	5.30	3.50	1.80	
Fovea2mm-peri2B	728	223	7.15	2.57	4.58	
	624	224	6.99	2.58	4.42	
Fovea2mm	2264	879	6.20	3.08	3.12	1.38

Table 3.3: Background-corrected fluorescence (*Cy3* and *Cy5*) and variance stabilized (*Vsn3* and *Vsn5*) data for S-antigen and nuclear pore complex interacting protein (*NPIP*).

Name	Symbol	Accession	Exp	Cy3	Cy5	Vsn3	Vsn5	Vsn3/Vsn5
Rhodopsin	RHO	NM_000539	1,2	396	5968	4.34	3.78	-0.54±0.60
			3,4	53	594	2.82	2.85	-0.04±0.60
			5,6	213	1758	3.08	4.01	-0.92±0.60
Long-wave cone opsin	OPN1LW	M13305	1,2	47	253	1.54	1.37	-0.12±0.89
			3,4	26	320	1.47	1.26	0.21±0.89
			5,6	177	142	2.69	1.48	1.21±0.89
Yes-associated 65	YAP65	X80507	1,2	32	189	1.01	1.06	-0.36±0.5
			3,4	349	968	4.60	3.30	1.31±0.5
			5,6	745	327	5.33	1.81	3.52±0.5
Histone deacetylase 9	HDAC9	AC004744	1,2	90	595	2.66	2.01	-0.07±0.65
			3,4	288	915	4.46	1.75	2.70±0.65
			5,6	681	199	5.27	1.90	3.37±0.65
SLT-ROBO GTPase act'ing	SRGAP2	AB032982	1,2	672	2222	4.03	3.87	0.88±0.88
			3,4	1750	6301	6.10	4.49	1.60±0.88
			5,6	1234	496	5.64	2.46	3.18±0.88
Hypothetical protein	FLJ	AK000110	1,2	750	3192	4.29	4.12	0.48±0.48
			3,4	705	4257	5.29	4.08	1.20±0.48
			5,6	1238	418	5.79	2.61	3.18±0.48
Nuc. pore comp. int'ing	NPIP	NM_006985	1,2	484	1868	3.67	3.6	0.28±0.28
			3,4	750	3161	5.32	4.24	1.08±0.28
			5,6	2264	879	6.20	3.08	3.12±0.28
cDNA from 6q22.1	6q	AL133101	1,2	896	2939	4.51	4.04	0.62±0.62
			3,4	1121	5100	5.70	4.33	1.37±0.62
			5,6	1212	494	5.82	2.73	3.09±0.62
cDNA from Xq13.1	X	AL110203	1,2	692	3595	4.25	4.09	0.56±0.56
			3,4	456	1839	4.89	3.63	1.26±0.56
			5,6	1014	389	5.66	2.57	3.09±0.56
β-homoCys CH3-t'ase	BHMT2	AK000008	1,2	3485	16485	5.9	5.67	0.59±0.59
			3,4	6946	27782	7.57	6.15	1.42±0.59
			5,6	13332	4995	8.22	5.16	3.06±0.59
GCN2 eIF2alpha kinase	GCN2	AB037759	1,2	1670	6990	5.03	4.91	0.71±0.71
			3,4	2306	9914	6.45	5.08	1.37±0.71
			5,6	4277	1636	6.95	3.92	3.03±0.71
Carboxylesterase-related	CRP	NM_016280	1,2	1221	5159	4.7	4.57	0.79±0.79
			3,4	2342	10425	6.47	5.10	1.37±0.79
			5,6	3432	1310	6.88	3.87	3.01±0.79
Retinitis pigmentosa 2	RP2	AJ007590	1,2	2015	9052	5.27	5.15	0.32±0.32
			3,4	3318	13635	6.86	5.60	1.26±0.32
			5,6	6912	2554	7.62	4.62	3.01±0.32
S-antigen (rod-arrestin)	SAG	X12453	1,2	1107	7351	4.98	4.64	-0.09±0.34
			3,4	36	282	2.41	2.09	0.32±0.34
			5,6	57	537	0.81	2.80	-2.00±0.34

Table 3.4: Summary of microarray data. Experiments 1&2 are controls, 3&4 are macular (4mm) *vs.* peripheral and 5&6 are foveo-macular (2mm) *vs.* peripheral retina. *Cy3* and *Cy5* are the mean signal values, for each background-corrected channel, from the two grouped duplicate experiments. The *Vsn3* and *Vsn5* columns are the variance stabilised expression values for each channel. *Vsn3:Vsn5* is the variance stabilized expression ratio; a positive value enumerates over-expression in the Cy3-labelled sample. The Cy3 channel represents macula or foveo-macula compared with peripheral retina in the non-control experiments.

in array manufacture prevented this from being an issue. Also, the SMA *lowess* algorithm did not seem to make a significant difference. Both the SMA *scale* algorithm (for scaling the array data so that each channel has the same median absolute deviation) and *variance stabilization* – applied globally to all the intensity values within an array – produced a marked reduction in the non-linearity of the control data.

Variance stabilization reduced the variability in low-expressed genes (see Figure 3.2 and Figure 3.3, lower graphs). It also produced more consistent results for several control genes, including the rod and red/green cone opsin (see Table 3.1 and Table 3.2) in which it showed the lowest co-efficient of variance for expression over these eight spots in slides 5 and 6. *Variance stabilization* was therefore chosen as the most appropriate algorithm with which to perform calibration before mining the data. *Variance stabilization* is a recent normalization algorithm that is freely available for research use and is designed to produce a constant signal-to-noise ratio over the entire data-set [Huber et al., 2002].

All non-control experiments except experiment 3 (4mm pooled macula *vs.* peripheral retina 1) showed a greater central bulge of mid-intensity differentially expressed genes than the control experiments (see Figure 3.3). The reason for this not being observed in experiment 3 could be that this particular peripheral retinal section was taken fairly close to the macular region, thus diminishing differences in expression. This underlines the need to perform more than one comparison. A handful of genes with high measured expression also appeared highly over-expressed in central retina, appearing as 'flares', especially in plots of experiments 4 and 6. Whilst some of the points in these 'flares' represent genes that were consistently over-expressed in central retina, others represent probes with variable expression data. These appear at a high mean expression level A as the mean is skewed by the higher expression value in the Cy3 channel. By calculating overall mean expression ratios using duplicate spots, by using data from two equivalent experiments, and by applying strict inclusion criteria for differentially expressed genes, it was possible to eliminate such artifacts from the analysis.

Like in all microarrays, expression data for individual probes showed some variation

both between slides (inter-array) and between duplicate spots on the same slide (intra-array). I defined the intra-array variation, for a particular gene, as the average deviation of the self-self duplicate (experiment 1) expression ratios from zero (see Tables 3.2 and 3.4. I defined the inter-array variation as the average deviation between two different samples of peripheral retina (experiment 2) from zero. I took the mean of these two variation figures to be a combined measure of variation for a particular gene. This gene-specific 'error value' was used to help interpret the mean of the four (duplicate) non-control experiments: foveo-macula (experiments 3&4) and fovea (experiments 5&6) *vs.* peripheral retina. I also calculated the average deviation of the expression ratios from the mean for these two sets of experiments.

Intra- was generally lower than inter-array variation, as expected, although both were less than 1 unit for most genes analysed. Whilst intra-array variation was likely due to differences in the spotted probe concentration, inter-array variation can be explained by differences in biological sample preparation and labelling. The combined variation was generally around ± 0.5, which on the log_2 scale is <1.5 fold. This is significantly less than the cut-off ratios of -2.0 (4-fold under-expression) and 3.0 (8-fold over-expression) used to choose genes for further study. Only one gene that was under-expressed in central retina (S-antigen) met the strict criteria for further study.

The average deviation of expression ratios from the mean in both groups of 2 non-control experiments was generally larger for genes with higher measured hybridization intensity. This may be due to differences in the two samples of peripheral retina being exaggerated in comparing them with central retina. This could perhaps have been minimized by applying *variance stabilization* simultaneously to all of the microarray slides instead of separately to each slide. As means were used, it is unlikely that this affected the final analysis. If definitive expression profiles of a large number of genes were sought in the foveo-macula, it is likely that a much larger amount of tissue would be needed to perform many more array experiments.

Variations in microarray data demonstrate the necessity to combine a number of ex-

periments to look for consistent differential expression. Similar principles are likely to apply to other methods of looking at differential gene expression, including SAGE. When this technique was used to look at differential expression in the macula by Sharon et al. [2002], 36 SAGE tags were reported to be significantly differentially expressed between two peripheral retinas (HPR1/2_SAGE). Only 24 SAGE tags were reported to be significantly over-represented in 'macula'[2] (HMAC2_SAGE) compared with peripheral retina (HPR2_SAGE).

I am often asked whether someone's 'favourite' gene is expressed more in central or peripheral retina, according to my microarray data. For many genes it is difficult to give a definitive answer as the variability in expression is too high relative to the expression ratio. This is explained by either the gene being expressed at a low level or that the probe did not work well enough. If more RNA is used then more probes will light up, although probes corresponding to higher-expressed genes will saturate and be non-informative in differential expression studies. The construction of arrays with non-optimal probe sequences could be overcome by designing slides with more than one probe for each gene. For example, in the case of both long-wave cone opsin and rhodopsin, it is likely that the probes were not optimal as their absolute fluorescence values were lower than those of many other genes (see Table 3.4), whereas we know that opsins represent a substantial portion of the retinal transcriptome.

Another factor in array design is the length of the probes. Whilst very short probes (25-mers or less) do not suffer from secondary structure formation they may be less specific than mid-length probes as a 25-mer can match more than one gene transcript at the lower hybridization temperatures that are necessary. Longer probes suffer from an increasing amount of secondary structure competing with the targets for binding and also an increasing amount of non-specificity. For example if whole cDNAs are arrayed, longer motifs may be similar enough to match more than one related gene. These limitations can often be overcome if enough experiments are performed, but it is necessary to weigh them up when choosing an array to use. The arrays I used were of 70-mers, which is near to

[2]One 6mm region of central retina was used, whereas the macula is defined as the central 4mm.

the optimal theoretical probe length. Cost is a major factor and the common Affymetrix 25-mer array platform was not available to me cheaply in Vancouver, where I procured the human macular resource. The Affymetrix platform is a one-colour non-fluorescent array. It would therefore also have been necessary to perform twice the number of experiments. More recent Affymetrix arrays have the advantage of using 5 different 25-mers for each gene, which goes a long way towards creating optimal probes.

3.5 Conclusions

I used microarray analysis to select a manageable number of candidate genes for further study. I performed multiple microarray experiments and compared several techniques of data analysis. I chose the most suitable method of normalization and mined the data by using a custom-built database. I chose candidate genes for further study using a set of strict selection criteria.

Chapter 4

Quantitative PCR

4.1 Introduction

The main alternatives to microarray analysis in the study of gene expression are reverse transcriptase PCR (RT-PCR) and Northern blotting. Northern blotting is a laborious procedure with low sensitivity [Bustin, 2000] that therefore requires a large amount of RNA for each gene studied. It also suffers similar limitations to microarrays because of its basis on hybridization: it is difficult to accurately compare expression between genes because of varying probe efficiencies. Quantitative PCR has now become widely used to validate expression data obtained from microarray experiments [Nadon and Shoemaker, 2002] because of its exquisite sensitivity and large dynamic range of up to 8 log units.

Quantification of PCR by measuring the amount of product at an end-point in the reaction is problematic because of the plateau that occurs as reaction components become limiting and PCR products compete for polymerase binding.

Competitive PCR [Becker-André and Hahlbrock, 1989] requires many extra reactions and optimisations. First, a mutated template (*competimer*) must be constructed that is amplifiable by the same primer pair as the query template, but can be distinguished from it by a unique restriction site or a size difference of tens of nucleotides. This competimer is then added at a known quantity to the test sample. PCR is performed until a measurable amount of product is obtained from both template and *competimer* in the exponential

phase of amplification.

The advent of 'real-time' PCR has brought quantitative PCR into the mainstream [Ginzinger, 2002; Walker, 2002]. Real-time PCR combines improvements in fluorescence chemistry with thermal cycler technology. A precise light source such as a laser is used to stimulate fluorescence, which is detected – via preferably high-fidelity optics – by a detector system such as a charge coupled device (CCD) or photomultiplier tube (PMT). This enables fluorescence measurement of PCR reactions on a cycle-by-cycle basis. A simple real-time PCR platform may consist of a UV lamp as the source, a CCD camera as the detector, and the inclusion of ethidium bromide (EtBr) in the PCR reactions.

The basis of quantitative PCR lies in the observation that *accumulation of fluorescence is proportional to accumulation of amplification products* [Higuchi et al., 1993]. Therefore, *the amount of fluorescence at cycle zero is proportional to the amount of template.* This initial fluorescence, and hence the amount of starting material, is too small to measure directly but can be calculated from the accumulation of fluorescence during subsequent PCR cycles.

Once blunt-ended copies of the template are produced in the second cycle of PCR[1], these products undergo doubling for each perfectly efficient additional cycle. The fluorescence-cycle amplification plot (Figure 4.1) is sigmoidal for real PCR data because the initial exponential increase in fluorescence reaches a plateau as the reaction efficiency decreases.

The exponential part of the amplification plot takes the form

$$f(C) = f(0) \times (1 + E)^C \tag{4.1}$$

where $f(C)$ is the fluorescence at cycle C and E is the reaction efficiency. E is a fraction between zero and perfect efficiency at a value of one. In quantitative PCR, the *threshold* is an arbitrary level of fluorescence within the exponential phase of amplification. The threshold cycle C_t is the PCR cycle during which the fluorescence reaches this threshold.

[1] I will not count these first 2 cycles as they complicate the analysis and can be ignored for purposes of relative quantification.

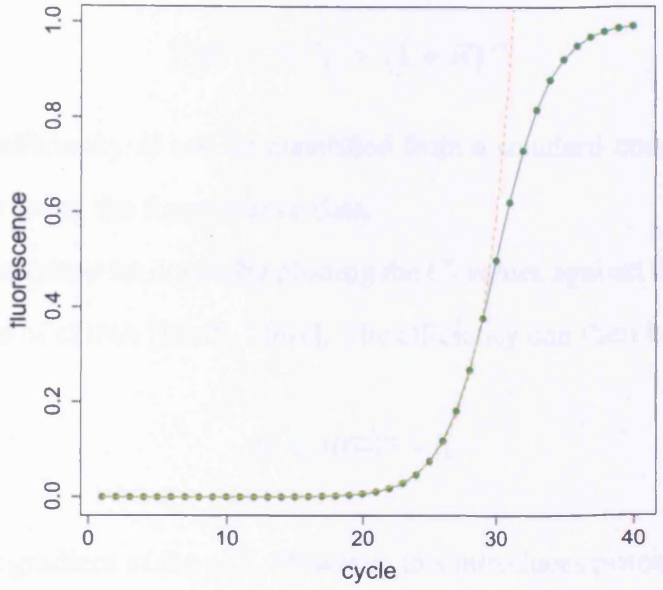


Figure 4.1: Graph to show qPCR data and an exponential model (orange line) of amplification. Fluorescence is shown using arbritrary units.

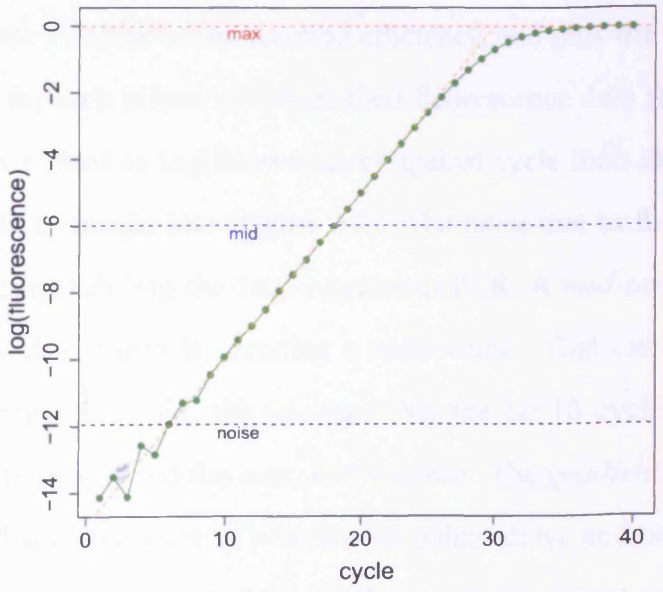


Figure 4.2: Amplification plot with log-transformed fluorescence values. The orange line is the gradient of the linear part of the the plot.

So to calculate the initial fluorescence $f(0)$ and thus the amount of template, Equation 4.1 can be re-written:

$$f(0) = f(C_t) \times (1 + E)^{-C_t}. \tag{4.2}$$

The reaction efficiency E can be quantified from a standard concentration series, by assumption, or by using the fluorescence data.

A standard curve may be drawn by plotting the C_t values against the log-concentration of a serial dilution of cDNA [Pfaffl, 2001]. The efficiency can then be calculated as

$$E = 10^{\frac{1}{slope}} - 1 \tag{4.3}$$

where *slope* is the gradient of the plot. However, this introduces potentially contaminating DNA and any pipetting or spectrophotometer errors will be amplified along with the DNA.

Commonly, the reaction efficiencies in quantitative PCR are assumed to be perfect ($E = 1$). This is referred to throughout the literature as the $2^{-\Delta\Delta C_t}$ method [Livak and Schmittgen, 2001]. Whilst this is a good approximation, it tends to exaggerate differential expression.

A more accurate measure of the reaction efficiency, and thus the amount of template, can be calculated for each primer pair from their fluorescence data [Peirson et al., 2003]. If amplification is plotted as log(fluorescence) against cycle then the exponential phase is transformed into a straight line (Figure 4.2). The noise due to fluorescence detection is particularly apparent during the 1st 10 cycles of PCR. A *mid*-point of the linear part of the plot can be determined by defining a *noise* value. This can be estimated as the minimum log(fluorescence) plus the *variance* over the 1st 10 cycles. The *mid*-point is halfway between the *noise* and the *max.* fluorescence. The *gradient* of the plot can then be calculated by linear regression of several data-points above and below the *mid* value.

The *gradient* of the linear part of the log(fluorescence) against cycle plot (Figure 4.2,

orange line) can be written as:

$$gradient = \frac{\log f(C) - \log f(0)}{C}. \tag{4.4}$$

The efficiency E of amplification can then calculated by re-arranging Equation 4.1 as:

$$(1 + E)^C = \frac{f(C)}{f(0)}. \tag{4.5}$$

Taking logs,

$$C \times \log(1 + E) = \log f(C) - \log f(0). \tag{4.6}$$

Dividing by C,

$$\log(1 + E) = \frac{\log f(C) - \log f(0)}{C} \tag{4.7}$$

which is equal to the *gradient*. Finally, by exponentiating and then subtracting 1,

$$E = 10^{gradient} - 1. \tag{4.8}$$

Data from qPCR, although much more accurate than Northern blotting, are often normalized using RNA quantification and/or a single 'housekeeping gene' as an internal control. Using RNA quantification alone is inappropriate because of the propagation of measuring errors and because rRNA, the major contributor to quantification, does not degrade at the same rate as all transcripts. The expression of so-called housekeeping genes is also variable. This is especially relevant when using post-mortem tissue [Tricarico et al., 2002]. Vandesompele et al. [2002] describe an approach, in which they recommend using at least three separate genes as a robust internal control for quantitative PCR. They found single control normalization error values that point to inherent noisy oscillations in expression levels of the control genes. This is corroborated by other large-scale expression studies including microarray data. [Warrington et al., 2000; Ross et al., 2000]

4.2 Objectives

After selecting twelve candidate genes that were differentially expressed in the foveo-macula by microarray analyses, my aim was to investigate these results further by using a complementary technique. My main objective was to confirm marked differential expression in a subset of these genes that would be suitable for further characterization. I planned to perform quantitative PCR using RNA from the human macular resource. I also planned to use the best possible technique for studying relative expression and normalization of the data, given that I had a limited supply of foveo-macular RNA. Finally I planned to investigate genes, using bioinformatics, that showed robust differential expression to gain useful information for protein work.

4.3 Results

Figure 4.3 shows that an intron-spanning fragment (see section 2.8) of rhodopsin was amplifiable as 405 and 239 base pair products from genomic DNA and peripheral retinal cDNA respectively. The smaller fragment, lacking the intron, was also amplifiable from a 1:10 dilution of the cDNA. There was no visible genomic fragment amplified from either cDNA sample.

Six out of six PCR products that I sequenced using the optimised qPCR primers (see Table 2.1) revealed amplification of the correct product.

I analysed the fluorescence data for each of the 12 test genes, red/green cone and rod opsin, and the 4 control genes using the DART system. An example of the analysis for red/green cone opsin is shown in Appendix B. A summary of the normalized expression for each non-control gene is shown in Table 4.1. A graph of the log-base-two expression ratios is also shown in Figure 4.4 where +1 is a two-fold increase and -1 a two-fold decrease in expression in foveo-macula:peripheral retina.

The amount of fluorescence due to SYBR-green is proportional to length of the DNA fragment amplified. Therefore by correcting for amplicon length, I calculated percent-

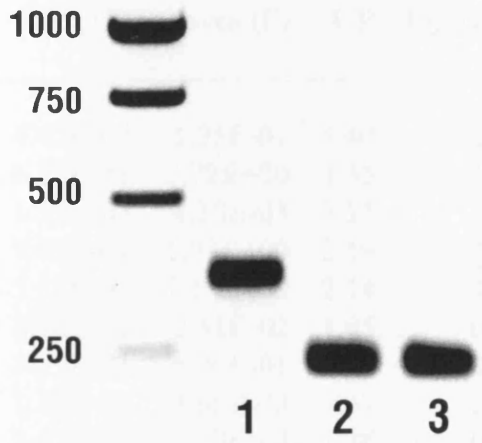


Figure 4.3: Agarose gel to show amplification of part of rhodopsin, using intron-spanning primers, from both genomic DNA and peripheral retinal cDNA subsequently used for qPCR. Lane 1 shows the 405bp product with genomic DNA. Lane 2 shows the 239bp product with peripheral retinal cDNA. Lane 3 contains the same product obtained from a 1:10 dilution of retinal cDNA.

age expression of selected genes relative to either rod or cone opsin (Table 4.1. NPIP, GCN2, and to a lesser extent X showed expression in the central retina approaching that of cone opsin, whilst SAG showed expression in the peripheral retina approaching that of rhodopsin.

4.4 Discussion

I used quantitative PCR as an experimental technique to analyse expression of twelve genes that were differentially expressed in the foveo-macula compared with peripheral retina according to microarray experiments (see chapter 3). I used a computational method [Peirson et al., 2003] to analyse the data and normalized it to four control genes.

The data show that the highest transcript levels were obtained for rod and cone opsin out of all the genes tested[2]. Red/green cone opsin was more abundant in the foveo-macula than rhodopsin, which was more abundant in the peripheral retina. The cone opsin message was also slightly more abundant in the foveo-macula than rhodopsin was in the peripheral retina. This may reflect the higher turnover of cones and their larger size.

[2]This excludes 18S rRNA, which represents over 10% of total RNA.

Symbol	Peri (P)	Fovea (F)	F/P	$\log_2$(F/P)	% opsin
X	4.42E-02	1.95E-01	4.40	2.14	7
OPN1LW	8.12E-01	2.72E+00	3.35	1.75	100
HDAC9	1.37E-03	4.30E-03	3.15	1.65	<1
NPIP	9.01E-01	1.97E+00	2.19	1.13	76
GCN2	5.37E-01	1.15E+00	2.14	1.10	34
SRGAP2	1.18E-02	2.31E-02	1.95	0.96	1
FLJ	3.19E-01	5.76E-01	1.81	0.85	*
6q	1.80E-03	2.65E-03	1.47	0.56	*
CRP	2.67E-04	3.63E-04	1.36	0.44	*
BHM2	8.39E-05	1.14E-04	1.35	0.44	*
RP2	3.16E-03	2.36E-03	0.75	-0.42	*
YAP	3.80E-03	2.49E-03	0.65	-0.61	*
RHO	1.90E+00	4.16E-01	0.22	-2.20	100
SAG	5.70E-01	1.06E-01	0.19	-2.43	34

Table 4.1: Summary of qPCR data. The first two columns show initial fluorescence $f(0)$ values for peripheral and central retina, normalized to the four internal control genes (18S, ARP, ACTB, UBC), which are proportional to mRNA copy numbers. The third column is the ratio of $f(0)$ values, the expression ratio, in central:peripheral retina. The fourth column is the base-two logarithm of this ratio. The final column is expression level as a percentage of the appropriate opsin expression.

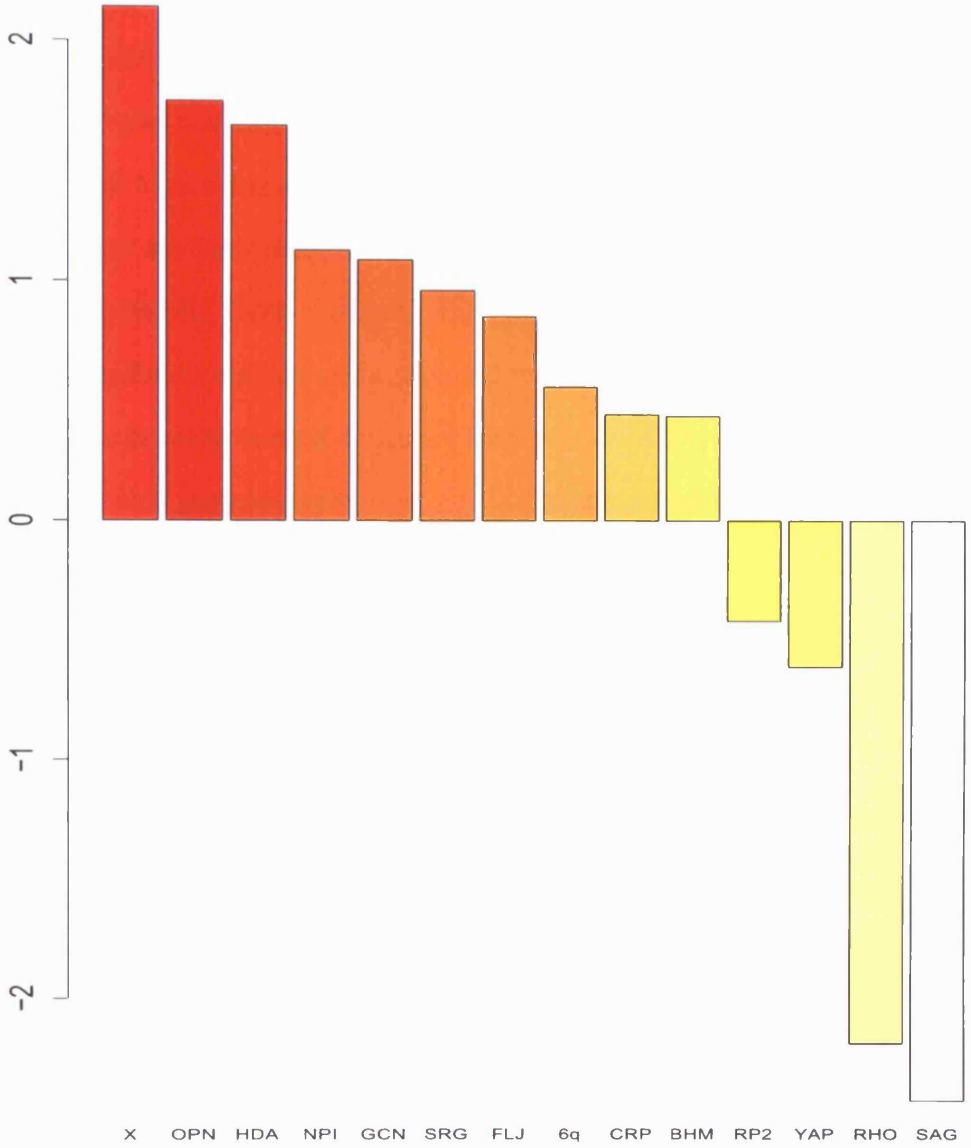


Figure 4.4: Graph to show quantitative PCR data comparing gene expression in pooled foveo-macular with pooled peripheral retinal RNA. Log-base-two expression ratios are shown. OPN=OPN1, HDA=HDAC9, NPI=NPIP, GCN=GCN2, SRG=SRGAP2.

Results for all but two out of the twelve test genes were in the same direction as the microarray expression data, with fold changes ranging from over four (for over-expressed genes) to less than one fifth (for under-expressed genes) in foveo-macula cf. peripheral retina. Because I used the amplification plot method to determine reaction efficiencies, the fold-changes are less likely to be exaggerated than with the commonly used $2^{-\Delta\Delta C_t}$ method. For example, there was only about a two-fold difference in expression of both opsins between foveo-macula and peripheral retina, whereas the true ratio is likely to be higher than this because central cones are large.

Because pooled samples were used, no statistical measure of variation could be calculated for the comparison of foveo-macula with peripheral retina. The peripheral samples were pooled because the small size of the foveo-macula necessitated pooling for RNA extraction (see chapter 1). Pooling 11 foveo-macular and 9 peripheral retina samples does means that the results are subject to less sample-sample variation than if single retinal samples had been compared, resulting in an increased level of statistical certainty. It was not possible to quantify the inter-sample level of significance due to the limited amount of tissue available. However, the co-efficient of variance for each triplicate was typically less than 10% (see Appendix B).

I also calculated the relative expression levels, compared to rod or cone opsin, for genes that were two-or-more fold over or under-expressed in foveo-macula (Table 4.1). This suggests that NPIP, GCN2, SAG, and to a lesser extent X, represent a relatively large proportion of the retinal transcriptome.

It is well recognized that protein expression is not always proportional to mRNA expression. A recent study by Tian et al. [2004] even suggests that less than half the variation in protein expression can be explained by variation in mRNA levels. It would be interesting to perform proteomic analyses for a large number of genes in order to choose previously un-investigated proteins that are differentially expressed in the macula. However, genes that are highly over-expressed at the mRNA level may be of importance despite modest differences in protein expression due to higher turnover.

The gene products of NPIP, GCN2, X, SRGAP2 and HDAC9 are as yet uncharacterized in the retina. My data provides evidence that they are over-expressed in the foveo-macula. Additionally, the messages for NPIP, GCN2 and X are expressed at levels that are likely to represent a significant proportion of the foveo-macular transcriptome (Table 4.1). Of these five genes, only SRGAP2 is as yet associated with a monogenic disease phenotype. However, NPIP, GCN2 and SRGAP2 are all located within possible regions of linkage suggested by a recent whole-genome screen of age-related maculopathy [Schick et al., 2003].

4.4.1 Differentially expressed genes

For discussion of HDAC9 and NPIP, please see chapter 5 and chapter 6 respectively.

GCN2 - Eukaryotic translation initiation factor two-alpha kinase 4 (Hs.261587)

GCN2 was highly expressed in human retina, its expression level being 36% of the red/-green cone opsin expression level in the foveo-macula. Homo sapiens eukaryotic translation initiation factor two-alpha kinase 4 is located at 15q13.3. It is a protein kinase that phosphorylates the alpha subunit of eukaryotic translation initiation factor 2 (eIF2α). Such protein kinases are activated in response to environmental stresses resulting in down-regulation of protein synthesis [Harding et al., 2000].

Human GCN2 is 88% identical to Mus musculus eIF2α kinase. It is also highly similar to GCN2 or similar genes in pig, rat, frog, drosophila, mosquito and Arabidopsis. ENSEMBL predicts 3 alternate proteins for human GCN2 (ENSG00000128829), the largest of which has 35 exons over 86kB coding for a 1473aa protein.

Human GCN2 contains many functional domains with similar catalytic elements to yeast GCN2. However, eukaryotic GCN2 proteins have distinct regulatory domains from yeast, allowing different signals to regulate translation. No diseases are associated with human GCN2 at this time.

Three SAGE tags were found that map uniquely to this reliable UniGene cluster and

were contained in a human retinal pigment epithelium SAGE library (HRPE1_SAGE); of these three tags, one was also found in the human macular SAGE library (HMAC2_SAGE) [Sharon et al., 2002] (ENSEMBL, NCBI: UniGene, OMIM, LocusLink, GEO).

X - predicted gene (Hs.138411)

This predicted gene was relatively highly expressed in human retina, its expression level being about 7% of the red/green cone opsin expression level in the foveo-macula. X is part of UniGene cluster Hs.138411 at Xq13.1. This cluster contains more than 20 ESTs or mRNAs. Gene prediction programs (GenScan, GenomeScan, EST2GENOME) predict a single exon gene from the mRNA and a peptide (XP_098999) of 126aa.

A unique SAGE tag for this UniGene cluster was found in a normal cerebellum (SAGE-normal-cerebellum) library. Also, a non-unique tag was highly represented in the human macular (HMAC2-SAGE), two peripheral retinal (HPR1/2-SAGE) and retinal pigment epithelium (HRPE1-SAGE) libraries (NCBI:UniGene, GEO, BLAST).

SRGAP2 - slit-robo Rho GTPase activating protein 2 (Hs.7977)

This gene is located cytogenetically at 3p25.3. SRGAP2 is also known as SRGAP3 or MEGAP and is part of the slit-robo Rho GTPase activating protein (SRGAP) gene family. The SRGAPs contain an FCH domain, a RHOGAP domain and an SH3 domain. SRGAP2 is highly expressed in brain and plays a critical role in the slit-robo signal transduction pathway [Endris et al., 2002].

Several SRGAP2 transcript isoforms have been described, at least 2 of which have been shown to represent functional GTPase-activating proteins (GAPs) by an in vitro GAP assay. Molecular analysis in a patient with a balanced *de novo* translocation t(X;3) (p11.2;p25), hypotonia, and severe mental retardation – features characteristic of 3p-syndrome – showed a translocation breakpoint on chromosome 3 interrupted the previously unknown gene termed MEGAP (mental disorder-associated GAP protein). The phenotype was suggested to be caused by mis-regulation of neuronal signal-transduction

machinery controlling the correct migration of neurons and their axonal connectivity. Haplo-insufficiency of MEGAP is proposed to lead to abnormal development of neuronal structures important for normal cognitive function [Endris et al., 2002] (NCBI: UniGene, OMIM, LocusLink).

SAG - S-antigen (Hs.32721)

S-antigen is located cytogenetically at 2q37.1. A soluble protein of 405 amino acids with a molecular weight of 45kD, SAG is implicated in the inactivation of the phototransduction cascade. The active form of rhodopsin (R*, meta-rhodopsin-II) is bound by S-antigen, especially after phosphorylation by rhodopsin kinase. The binding of SAG blocks interaction of the G-protein transducin thus inactivating the visual cascade [Fain et al., 2001].

Oguchi disease (MIM:258100) is a recessively inherited form of stationary night blindness. Patients with this disease show a distinctive golden-brown colour of the fundus that occurs as the retina adapts to light, called the Mizuo phenomenon. In many cases Oguchi disease is due to mutation in the SAG gene [Fuchs et al., 1995]. Adaptation of rod photoreceptors to light is extremely retarded in Oguchi disease, while that of cones appears to proceed normally.

Whilst some studies report the expression of the SAG protein only in rod photoreceptors [Reid et al., 1987; McKechnie et al., 1986], there is evidence that human rods and cones contain antigenically distinct S-antigens [Nork et al., 1993], and that mouse and rat also express S-antigen in both rods and cones [Mirshahi et al., 1994].

No unique SAGE tags were found for this gene. However, a non-unique tag was found to be well represented in only human peripheral retina (HPR1/2-SAGE) and macular (HMAC2-SAGE), with the representation being about two-fold greater in the peripheral retinal libraries (NCBI: UniGene, OMIM, GEO).

Although s-antigen is often referred to as rod-arrestin, there is evidence that it is also expressed in cones [Nork et al., 1993; Mirshahi et al., 1994]. My microarray and quantitative PCR data together provide strong evidence for over-expression of the s-antigen

message in the peripheral retina, perhaps by around five-fold. This is corroborated by the clinical picture of Oguchi disease, in which central vision is normal.

4.4.2 Proteins

I went on to study HDAC9 and NPIP in some detail, as outlined in the next two chapters. I initially chose to study HDAC9, NPIP and GCN2 as the three genes most highly over-expressed in central retina (excluding X) according to quantitative PCR. X was excluded as it is a one exon predicted gene so despite plenty of hits in dbEST, the possibility that X did not represent a 'real' gene could not be excluded. The number of proteins studied was limited by the resources available to obtain good antibodies.

4.5 Conclusions

The main aim of the qPCR experiments was to confirm the expression data obtained from microarray analysis. It is encouraging that both opsins and ten out of the twelve test genes showed expression values that agreed with those obtained from microarray analysis. I used effective methods of analysis and normalization given the limited amount of foveo-macular RNA available and also determined the expression levels of the genes studied relative to opsin. The protein products of all the genes with confirmed over-expression in the macula (except cone opsin) are as yet uncharacterized in the retina. They are suitable targets for the investigation of protein expression in the macula. These genes are likely to play a role in the mechanism of precise vision and also represent good candidates for macular disease.

Chapter 5

Histone deacetylase 9

5.1 Introduction

Chromatin is a highly specialized structure composed of tightly compacted chromosomal DNA. Transcription is controlled by a host of protein complexes that continuously pack and unpack chromosomal DNA. The *nucleosome* is the basic structural unit of eukaryotic nuclear *chromosomes*. A *nucleosome* consists of 1.75 superhelical turns of chromosomal DNA wrapped around a histone octamer (two molecules each of the four core histones H2A, H2B, H3 and H4) via a single molecule of the linker histone H1. *Histones* are simple, *basic* proteins that are soluble in water.

One of the mechanisms that unwinds *nucleosomes* from inaccessible, tightly packed units into accessible particles involves the acetylation and deacetylation of the histone proteins comprising its core. Acetylation of histones confers accessibility of the DNA template. It is a major factor in regulating chromatin structural dynamics during transcription [Lee et al., 1993; Bauer et al., 1994].

Histone deacetylases (HDACs) are essentially chromatin remodelling factors that deacetylate histones and thus act as transcriptional repressors. HDACs are classified by their sequence homology to the yeast HDACs. Class I HDACs are related to *Rpd3* while members of class II are larger molecules (100+ kDa) that resemble *Hda1p*.

HDAC9 is part of the class II histone deacetylase gene family and is orthologous

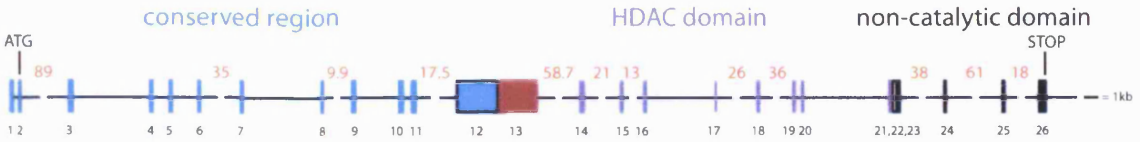


Figure 5.1: Schematic diagram of HDAC9. Exon (box) numbers are shown underneath in black. Where introns (lines) are not shown to scale, the intronic size appears in red above.

to the corresponding Xenopus and mouse genes. It was originally identified by Nagase et al. [1998] while looking for cDNAs that encoded large proteins expressed in the human brain. Their RT-PCR analysis detected highest expression in brain, lower expression in heart and smooth muscle, and little or no expression in other tissues. Zhou et al. [2001] found 6 isoforms of HDAC9 and detected preferential expression of the longer isoform, by RT-PCR, in brain, heart, and pancreas. They also determined that the full-length human gene contains 23 (out of 26) coding exons spanning 458 kb and explained the variable exon usage of different HDAC9 isoforms. Petrie et al. [2003] showed that HDAC9 encodes multiple, functionally distinct, and differentially expressed protein isoforms. For example, full-length HDAC9 was localized to the nucleus whereas the isoform lacking exon 7 was expressed in the cytoplasm.

Full-length HDAC9 (see Figure 5.1) consists of an N-terminal region that is conserved in HDAC 4, 5 and 7, an HDAC catalytic domain in the C-half, and a C-terminal non-catalytic domain. Histone deacetylase 9 (HDAC9) is located cytogenetically at 7p21.1. HDAC9's huge genomic size ($\approx$ 500 kb) and the degree to which it is regulated is unprecedented among other HDACs. This may indicate a wider role for HDAC9 than just histone modification [Petrie et al., 2003].

It has recently been suggested that HDAC9 may be jointly (with TGFβ2) responsible for possible *digenic* inheritance of Peter's Anomaly. This is a disease of the anterior segment of the eye thought to be the result of abnormal migration or function of neural crest cells [David et al., 2003]. Also, mice lacking HDAC9 are known to be super-sensitive to stress signals [Antos et al., 2003; Zhang et al., 2002]. This is consistent with the hypothesis that histone deacetylases are involved in the cellular response to stress.

5.2 Objectives

The data I obtained from microarray (chapter 3) and qPCR (chapter 4) analyses showed that HDAC9 is highly differentially expressed in the human macula. My aim was to study the expression of HDAC9 protein. I planned to obtain an antibody to HDAC9 that was useful and specific in both Western blotting and immunohistochemistry. I aimed to use the antibody for Western blotting to detect HDAC9 in cultured cell lines and in retinal extract. I planned to test the specificity of the antibody for immunohistochemistry in cultured cells and then go on to use it in human retinal sections to elucidate the localization and possible function for HDAC9 in the macula. By studying HDAC9 protein in the retina, I also set out to assess it as a possible candidate for macular disease.

5.3 Results

I detected a single band by Western blotting of REH cells, Chinese hamster ovary (CHO) cells, and 293T cells with a 1:500 dilution of rabbit polyclonal anti-HDAC9$_{C-term}$ antibody (Figure 5.2). The REH cell band was very dark, thick and slightly smeared, even after a short exposure of 15 seconds. I was not able to detect full-length HDAC9 in human retina.

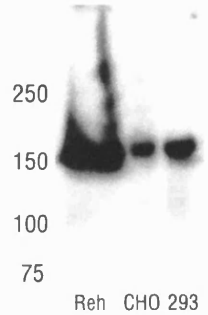


Figure 5.2: Western blot showing expression of full-length HDAC9 in mammalian cell lines using a rabbit polyclonal anti-HDAC9$_{C-term}$ antibody.

I was able to localize full-length HDAC9 in adherent cells (SK-N-SH) using the anti-HDAC9$_{C-term}$ polyclonal antibody (Figure 5.3). Using confocal microscopy I was able to focus on several cells in more than one plane. I found that anti-HDAC9$_{C-term}$ was

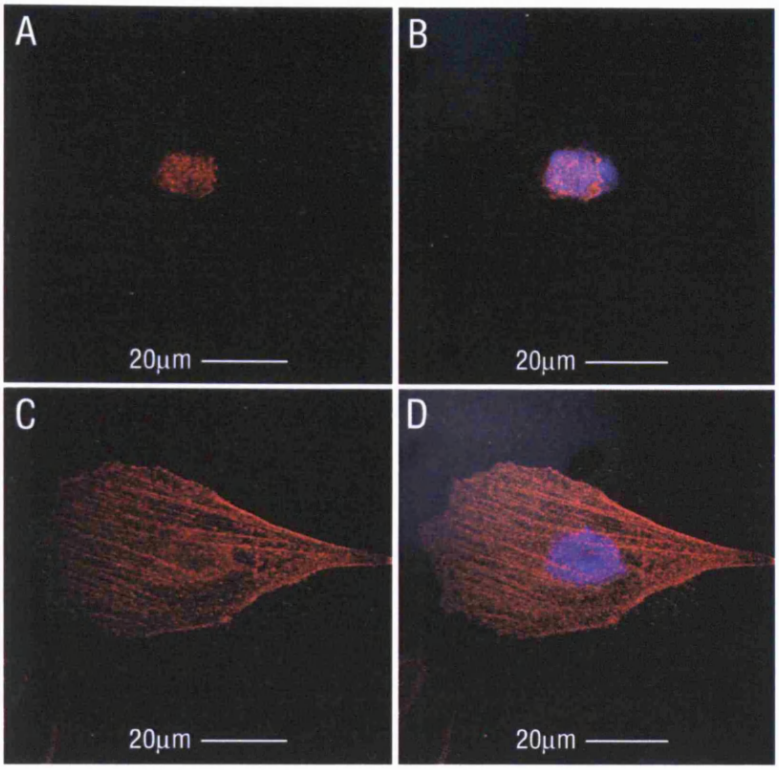


Figure 5.3: Confocal microscope images of cultured human neuroblastoma (SK-N-SH) cells. Untransfected cells were immunostained with 1:250 rabbit anti-hdac9 serum and a Cy3 secondary antibody and counter-stained with DAPI. **A** - red fluorescence showing HDAC9 expression in the nucleus. **B** - merged image showing DIC, DAPI staining of the nucleus and red HDAC9 fluorescence. **C** - the same cell in a different plane shows a filamentous staining pattern with anti-HDAC9$_{C-term}$. **D** - merged DIC, DAPI and anti-HDAC9 image.

not only localized to the nucleus, as evident by its co-localization with DAPI, but also to filamentous structures in a different plane of the cell.

I was also able to localize full-length HDAC9 in the human fovea using cryo-sections and the same titre of anti-HDAC9$_{C-term}$ (Figure 5.4). I saw fluorescence in many cell nuclei in the outer and inner nuclear layers, and in several nuclei of the ganglion cell layer. This was re-enforced by co-localization of the DAPI fluorescence. I also saw marked anti-HDAC9$_{C-term}$ fluorescence in the photoreceptor layer, which consisted mostly of cones in the cryo-sections that were available. This was strongest in the photoreceptor outer segments. There was also fluorescence of the inner limiting membrane but very little anti-HDAC9$_{C-term}$ signal in the plexiform layers.

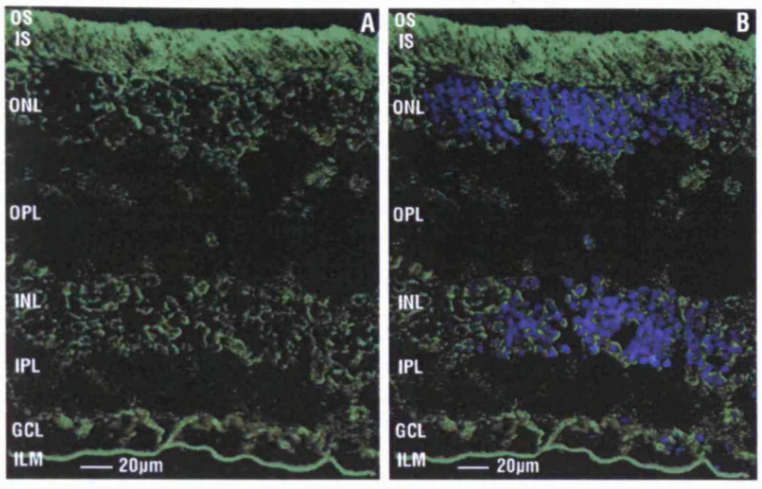


Figure 5.4: Full-length HDAC9 expression in human fovea (A). Anti-HDAC9$_{C-term}$ (green) fluorescence is also shown against DAPI (blue) nuclear staining (B).

5.4 Discussion

I was able to show expression of full-length HDAC9 protein (≈160 kDa) in both non-primate (CHO) and primate (293T, REH) cell lines (Figure 5.2). It was highly expressed in the REH cells that I received from Arthur Zelent's lab (which I successfully cultured) as previously reported [Petrie et al., 2003]. I was not able to detect the full-length HDAC9 by Western blotting in the human retina sample I used. This may have been partly because I received the particular retina I used more than 48 hours after enucleation. The retina is also very difficult to homogenize so this may have contributed to a low level of protein. My quantitative PCR data (chapter 4 Table 4.1) suggested that the HDAC9 gene was only expressed at a low level in the retina. However, gene expression does not necessarily correlate with protein expression [Tian et al., 2004]. HDAC9 may be important in retinal function or turnover even if the protein is not expressed at a high level. For example, the X-linked retinitis pigmentosa protein RP2 is ubiquitously expressed at relatively low levels (0.01% of total protein) [Chapple et al., 2000].

I localized full-length HDAC9 to the nucleus of human neuroblastoma (SK-N-SH) cells (Figure 5.3). Full-length HDAC9 has been previously localized to the nucleus by transiently transfecting simian fibroblast (COS-7) cells with an HDAC9-FLAG construct followed by immunofluorescence with an anti-FLAG antibody [Petrie et al., 2003]. It

is logical that a histone deacetylase protein is expressed in the nucleus, although there are many members of the gene family, some of which may have alternative functional specializations. I also detected full-length HDAC9 in filamentous structures that looked very much like part of the cytoskeleton. Interestingly, histone deacetylase 6 (HDAC6) has been recently found to act as a molecular motor in the cytoskeleton [Kopito, 2003].

I was also able to localize full-length HDAC9 in human fovea. That I was able to detect HDAC9 by immunofluorescence of the fovea but not by Western blotting of whole retina may reflect a higher level of protein preservation in the cryo-sections I received from Canada and also higher expression of HDAC9 in the fovea. Full-length HDAC9 is expressed in the outer and inner nuclear layers, the ganglion cell nuclei and the inner limiting membrane (Figure 5.4). However, the expression is greatest in the photoreceptors, especially in the outer segments.

Histone deacetylation is correlated with the transcriptional repression and silencing of genes. Genetic repression may have a protective role in neuronal aging and degeneration; histone acetylation induces a stress response that leads to apoptosis of neuronal cells [Salminen et al., 1998]. This may be especially important in the macula, which is likely to have a higher turnover of proteins than the peripheral retina.[1]

The presence of full-length HDAC9 in outer segments suggests that it may have a function related to the photo-transduction cascade or its maintenance. Mutations in many of the proteins expressed in photoreceptor outer segments are responsible for retinal degenerative disease. Interestingly, there is also a retina-specific promoter of HDAC9 in mice [K Petrie, personal communication]. *Histone* is known to activate photoreceptor outer segment cyclic nucleotide phosphodiesterase (PDE) [Hurwitz et al., 1984a,b]. It is possible that HDAC9 has a modulatory effect on phototransduction via the deacetylation of histone. HDAC9 could also be involved in the transport of molecular cargo to the outer segment along microtubules.

The more than three-fold over-expression of HDAC9 in the foveo-macula compared

[1]The macula comprises <1.5% of the retina yet it contains ≈8.5% of the cone photoreceptors, ≈3.5% of the rod photoreceptors, and ≈60% of the ganglion cells of the retina [Sharon et al., 2002].

with peripheral retina (see chapter 3 and chapter 4) also suggests that it is enriched in cone photoreceptors. Mutations in HDAC9 may therefore be responsible for macular, cone, or cone-rod dystrophy. HDAC9 may also play an important part in age-related macular degeneration. Autosomal dominant cystoid macular dystrophy has been mapped to the same chromosomal location as HDAC9 (7p21-15), although it is thought to be due leaking peri-macular capillaries [Kremer et al., 1994].

5.5 Conclusions

I was able to obtain an antibody to HDAC9 that was useful and specific in both Western blotting and immunohistochemistry. As well as nuclear localization, I showed a novel cytoskeletal localization of endogenous HDAC9 in human neuroblastoma cells, similar to HDAC6. Although I was not able to detect HDAC9 in retinal extract by Western blotting, I did show that HDAC9 is expressed in the human macula by immunofluorescence. These observations suggest a novel role for HDAC9 in the macula. This is the first implication, to my knowledge, of a histone deacetylase playing a part in retina-specific biology. In summary, I have made some headway towards eliciting the function of HDAC9 in the macula and demonstrated that it is an interesting candidate for macular disease.

Chapter 6

Morpheus

The *morpheus* gene family found within a 20kb duplicated segment – low-copy repeat sequence 'a' (LCR16a) – throughout 15Mb of human chromosome 16p [Johnson et al., 2001]. It is the result of *adaptive evolution* during the emergence of humans and African apes.

Adaptive evolution was originally defined by Darwin [1859] when he made the observation that island finches had a diversity of mouth parts that was inexplicable by usual trends in variation. The original population of birds had not inhabited the island for long yet they had developed a huge diversity of bill morphologies. Darwin concluded that there was *selective pressure* on the birds to adapt very quickly to many small niches on the island. This facilitated the rapid *adaptive evolution* of their mouth parts – some for digging in the sand, some for nut-cracking, some for scavenging under the bark and so on.

The concept of *adaptation* is strongly linked to function. A trait is considered to be *adaptive* if it confers a function that is beneficial to the organism. However, this does not imply any theory about how the *adaptive* trait originated.

Darwin's notion of *natural selection* is that evolution of a new *adaptive* trait can be explained if there is:

- A process by which a new trait can appear in a population.

- A biological mechanism for heredity.
- A "struggle for existence" arising from larger reproductive capacity than the environment can support.
- A benefit in terms of survival[1] of the trait in question.

So in the Darwinian sense, the term *adaptation does* imply how the trait originated. In Mendelian terms, *adaptation* benefits an individual by representation of its genes in the next generation. As Dobzhansky [1956] put it:

> *Selection favors genotypes the carriers of which transmit their genes to succeeding generations more efficiently than do the carriers of other genotypes.*

However, it is important to note that not all traits that are *adaptive* in the Darwinian sense are the result of *natural selection*. Although a trait is *adaptive*, it has not necessarily been 'selected for'. For example, a trait can arise in a species that is the result of random *fixation* of a *selectively neutral* allele (*genetic drift*). Later, environmental changes lead to the trait becoming *adaptive*.[2] An *adaptive* trait can also result from changes within the life of an individual in response to the environment, such as acclimatisation to altitude, that does not result in a change in gene frequency. In humans, a dramatic form of *adaptation* is illustrated by cultural differences. Cultural traits often yield Darwinian *fitness* that is not due to genetic differences.

Negative selection decreases the prevalence of traits that diminish an individual's *fitness*. Conversely, *positive selection* is a phenomenon whereby there is *selective pressure* favouring *adaptation*. *Negative selection* can be seen as a process of editing genetic change so that only a small number of mutations are retained in a population. With *positive selection*, the retention of mutations is much closer to the rate of mutation. Since the advent of molecular techniques, a common test for *positive selection* is to compare the

[1] *Sexual selection* can be defined by the same criteria as *natural selection* except that the trait is beneficial in terms of reproduction rather than survival.

[2] The Dykhuizen–Hartl effect [Dykhuizen and Hartl, 1980].

rates of synonymous to non-synonymous nucleotide change K_a/K_s [Lee, 1997]. A ratio significantly greater than 1 is taken as evidence for *positive selection.*

The evolutionary significance of *gene duplication* was first recognized by Haldane [1932] and Müller et al. [1935], who suggested that a redundant duplicate may acquire divergent mutations and eventually emerge as a new gene. *Gene duplication* was first observed by Bridges [1936] at the Drosophila *Bar* locus. *Gene duplication*, followed by functional divergence, is considered to be one of the most important mechanisms for the evolution of novel gene function [Ohno, 1970]. However, it has been controversial whether functional divergence (and hence *adaptive evolution*) occurs by *positive selection* [Goodman et al., 1975], or by random fixation of neutral mutations[3] that later become *adaptive* due to environmental changes [Dykhuizen and Hartl, 1980; Kimura, 1983]. A recent model for the evolution of new proteins is that a period of *gene sharing* [Piatigorsky et al., 1988] usually precedes the evolution of functionally distinct proteins. *Gene duplication* then allows each daughter gene to specialize for one of the functions of the ancestral gene [Hughes, 1994].

The principal molecular mechanism responsible for gene duplication is unequal crossing over [Grauer and Li, 2000]. This results in mis-aligned sequences giving rise to a tandemly duplicated region on one chromosome and a complementary deletion on the other. Unequal crossing over is facilitated by the presence of repeated sequences. Once a duplication event has occurred, the process of gene duplication can accelerate because the chance of unequal crossing over increases with the number of duplicated copies.

Many other mechanisms also play important roles in molecular evolution. For example, regional duplications such as gene elongation and domain duplication [Grauer and Li, 2000]. Gene conversion is an important mechanism that does not involve a duplication event. It is a type of nonreciprocal recombination in which the recipient strand of DNA receives information from another strand. The transfer occurs by DNA repair of a damaged allele using the donor strand as the template. The donor can be either the

[3]The neutral theory of evolution [Kimura, 1968, 1983] and subsequent molecular-clock hypothesis [Zuckerland and Pauling, 1965] are now thought to be too simple to entirely explain adaptation, although they provide a stable background against which to test *positive selection* at a molecular level.

sister chromatid or homologous chromosome during meiosis or a duplicate gene during mitosis. Both gene duplication [Dulai et al., 1999] and gene conversion have been shown to be important mechanisms in the evolution of red and green cone opsin in Old World monkeys [Ibbotson et al., 1992] and humans [Winderickx et al., 1993].

The dispersal of LCR16a was accompanied by considerable chromosome map and copy number variation among hominoids, as revealed by fluorescence *in situ* hybridization (FISH) [Johnson et al., 2001]. The estimated copy number in orangutans, gorillas, humans and chimpanzees is 9, 17, 15 and 25-30 respectively. By calculating the K_a/K_s ratios of putative protein encoding exons, Johnson et al. revealed the most extreme case of positive selection. Between Old World Monkeys and humans the ratio in exon 2 was 13.0. The major episode of enhanced amino acid replacement was calculated to occur after the separation of human and great ape from the orangutan lineage. *Positive selection* also continued after separation of the human and chimpanzee lineages. Johnson et al. did not find any significant sequence similarity to the *morpheus* gene family in non-primate species, which is likely because of the rapidity of its evolution. Interestingly, it is only primates that have a macula.

Johnson et al. found transcripts in dbEST (`ncbi.nlm.nih.gov`) for 6 out of 15 genes in the *morpheus* family. Amino acid sequence comparison of two full-length transcripts showed 81% identity. In sharp contrast, the corresponding introns were 98% identical. This is a hallmark of genes undergoing adaptive evolution [Vacquier et al., 1997]. One of these transcripts, nuclear pore complex interacting protein (NPIP), had been independently isolated by a two-hybrid assay in an effort to identify components that interact with the human SWI/SNF transcriptional activator complex. Johnson et al. predicted the presence of one transmembrane domain flanked by α-helical secondary structure. Expression of a green fluorescent protein (GFP) fusion construct showed localization to the nuclear envelope and co-localization with nucleoporin (p62). Whilst RT-PCR showed a broad distribution of the *morpheus* gene family in human tissues, no specific localization or function of one of its members, such as NPIP, has been found *in vivo*.

6.1 Objectives

As well as being consistently over-expressed in macula compared with peripheral retina, I found NPIP (and/or other very similar members of the *morpheus* gene family) to be very highly expressed in human retina at the mRNA level (see chapter 3 and chapter 4). My aim was to study the expression of proteins in the *morpheus* family. I planned to obtain an antibody to NPIP that was useful and specific in both Western blotting and immunohistochemistry. I aimed to use Western blotting to detect *morpheus* in the retina, in cultured cells to test the specificity of the antibody for immunohistochemistry, and in human retinal sections to elucidate the localization and possible function for *morpheus* in the macula. By studying proteins in the *morpheus* family in the retina, I also set out to assess them as possible candidates for macular disease.

6.2 Results

The NPIP protein AAD34394, a translation of the reference cDNA AF132984 aligned with a number of NPIP-like sequences from the database as well as predicted proteins obtained using the Genomescan program (see Appendix B). The translation of the amplicon used in qPCR lined up very well with these sequences. All but one of these putative proteins terminated at the same place as the reference NPIP sequence. However, the proteins seemed to begin in various places. The C-terminus of NPIP was chosen to raise an antibody against.

ELISA revealed that both of the two antisera produced bound to the NPIP peptide with increasing affinity as the titre of antibody was increased. Control serum did not bind to the NPIP peptide, nor did either anti-NPIP antibody bind to a control peptide.

I detected a number of different bands by Western blotting of primate cell lines using each antiserum, anti$_a$-NPIP and anti$_b$-NPIP, with anti$_b$-NPIP yielding less bands although the blots were not 100% specific with either antiserum under the conditions used.

I successfully cloned several NPIP cDNAs into pGEM-T Easy from both brain and

retina. All clones were correctly sized on a gel after digestion and sequenced before sub-cloning (see Appendix B for examples of sequence data.

I found two novel cDNA sequences in brain and one in retina. One brain sequence (NPIP-brain1, see Appendix B Figure B.6) was very similar to AF132984, except that there are an extra 47 bases within exon 8. Despite changing the reading frame it results in an almost identical protein to the 350aa AAD34394 with an additional 19 amino acids in the C-half (the only 2 changes are in the N-half, see Appendix B Figure B.5).

The second cDNA sequence (NPIP-brain2) I found was identical in more than one clone. Alignment with several genomic sequences from the *morpheus* gene family revealed a longer exon 1 using an alternative splice donor site that followed the GT-rule (Figure B.6). This would result in a very premature STOP codon within the novel exon 1A. However, the longest open reading frame of this cDNA codes for a very similar protein that is missing 123aa of the N-terminus (see Figure B.7).

I also cloned several cDNAs of the same length as NPIP-brain1 and NPIP-brain2 from foveo-macula and peripheral retina but did not obtain novel sequence that was the same in two or more clones.

I attempted to align and orientate the above transcripts with genomic DNA from the database, in order to find out from which part of chromosome 16p they originate, by exhaustive BLAST and BLAT (`genome.ucsc.edu`) searching and manual analysis of the hits to check the exons order and adherence to the AT-GT splice site rule. Each of the transcripts produced thousands of incomplete alignments in at least fifteen different genomic locations. However, none of these alignments produced an exact match for any of the transcripts when they were analysed by hand for exon order and the conservative splicing rule.

I was able to detect expressed NPIP-GST fusion protein by Western blotting using $anti_b$-NPIP primary, goat anti-rabbit secondary antibody, and ECL. I was subsequently to identify NPIP in both peripheral and whole retinal extract (Figure 6.1) with a primary antibody titre of 1:250. The two bands of $\approx$25 and $\approx$40 kDa correspond to size of the

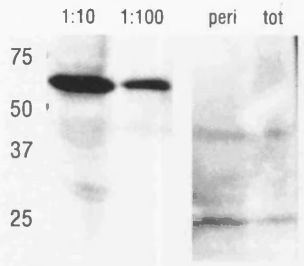


Figure 6.1: Western blot to show NPIP in human retina using 1:250 anti$_b$-NPIP rabbit polyclonal antibody. The first two lanes contain NPIP-GST and the last two contain peripheral and whole retinal extracts respectively.

putative NPIP isoforms found in brain and retinal cDNA.

I detected specific fluorescence with 1:200 anti$_b$-NPIP and a Cy5 donkey anti-rabbit secondary antibody in un-transfected SK-N-SH cells by using fixation in methanol for 15 minutes at -20°C (Figure 6.2). Staining was seen in the region of the nuclear envelope. I also expressed separate GFP constructs containing NPIP-brain1 and NPIP-brain2 in SK-N-SH cells (see Figure 6.3). NPIP-brain2 definitely did not localize to the nuclear pore – it was within the nucleus and in nuclear inclusion bodies. NPIP-brain1 localized to the nuclear pore (I also co-localized it with anti$_b$-NPIP) although imaging was difficult as the SK-N-SH cells became very rounded after transfection.

I was able to stain human macular cryo-sections under the same conditions. Confocal microscopy revealed staining of a sub-population of photoreceptors that looked more like cones than rods. However, the secondary antibody appeared speckled at high magnification. I changed the secondary antibody to Cy3 and used 1:150 Cy5-streptavidin with 1:100 biotinylated peanut agglutinin (PNA) for a co-localization study in human fovea Figure 6.4. I examined three sections of fovea and acquired many images at low and high power. I was also able to acquire several stacks of images where I focused the laser scanning microscope in several parallel planes to gain a 3-dimensional impression of localization. This revealed very specific staining in cone outer segments. There was also very close co-localization of PNA, which stains cone sheaths. In animations of the image stacks, it looked like the anti$_b$-NPIP staining was just inside the cone sheaths.

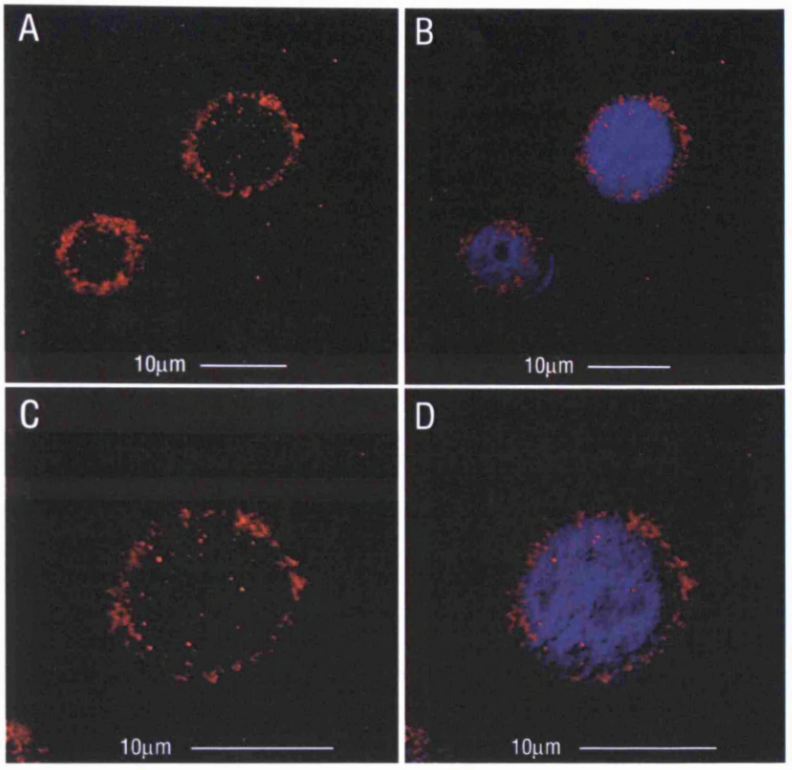


Figure 6.2: Confocal microscope images of cultured human neuroblastoma (SK-N-SH) cells. Untransfected cells were immuno-stained with 1:200 rabbit anti-NPIP serum and a Cy5 secondary antibody and counter-stained with DAPI. **A** - red NPIP fluorescence at medium magnification showing expression in the vicinity of the nuclear envelope. **B** - merged DIC (grey), DAPI (blue) and NPIP (red) image at medium magnification. **C** - Red NPIP fluorescence at high magnification. **D** - Merged DIC (grey), DAPI (blue) and NPIP (red) image at high magnification.

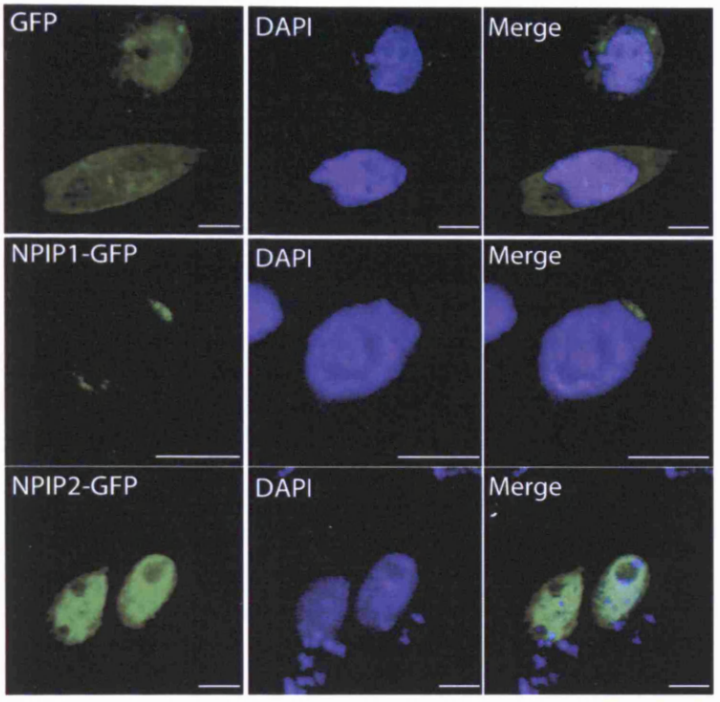


Figure 6.3: Confocal microscope images of cultured human neuroblastoma cells transfected with two different NPIP GFP constructs. Each row shows the same cells imaged under the same conditions for GFP (1st column) and DAPI (2nd column) and a merged image (3rd column). The first row shows cells transfected with vector containing a GFP construct. The 2nd and 3rd rows show cells transfected with vector containing GFP-NPIP-brain1 (**NPIP1**) and GFP-NPIP-brain2 (**NPIP2**) respectively. The white bars in each sub-figure represent a distance of ten microns.

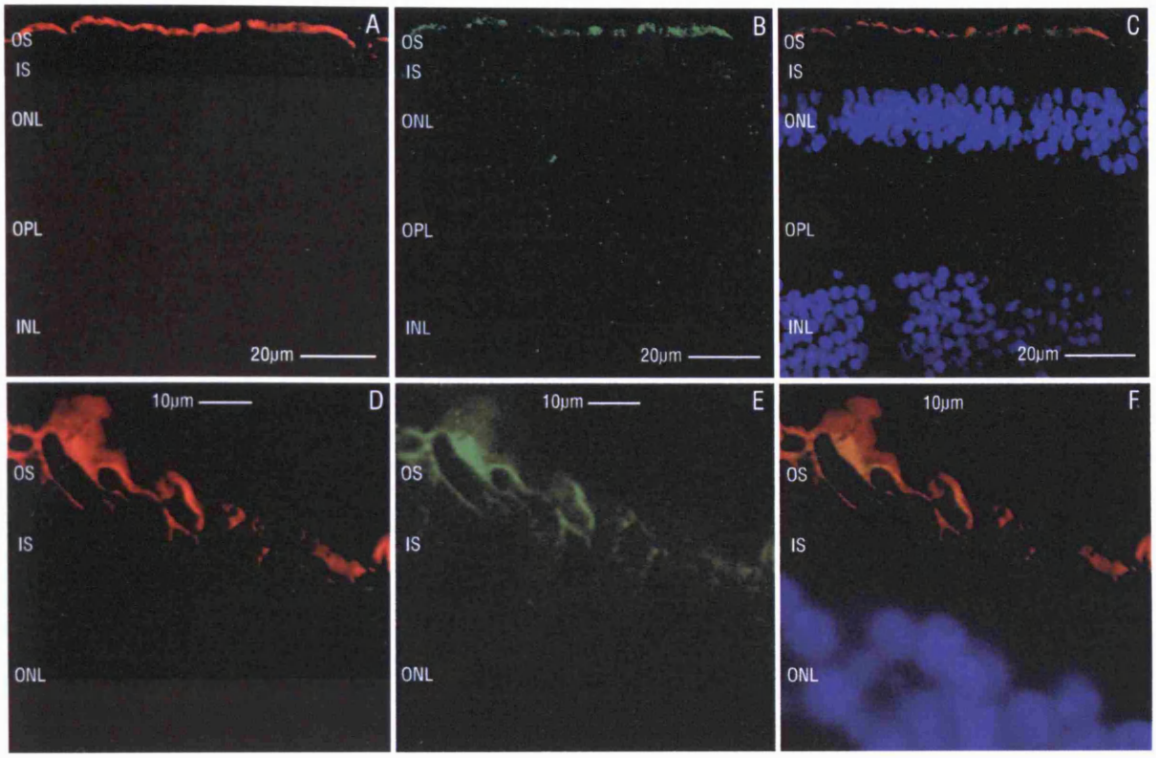


Figure 6.4: Confocal microscopy showing anti$_b$-NPIP and PNA staining in human fovea. At low magnification (red) peanut agglutinin (PNA) readily stains cone outer segments (A). Many outer segments also stain with (green) anti$_b$-NPIP (B,C). At high magnification the very close correspondence between PNA and anti$_b$-NPIP can be seen (D-F).

6.3 Discussion

As well as being consistently over-expressed in macula compared with peripheral retina, NPIP was found by quantitative PCR to be highly expressed in human retina (see chapter 4). Its expression level is 80% of the red/green cone opsin expression level in the foveo-macula. This PCR was very specific to members of the *morpheus* gene family as two long primers were used at a highly stringent temperature. However, it is not certain exactly which member of the *morpheus* gene family was picked up by the gene expression studies. There is a high degree of similarity between most exons amongst the gene family members [Johnson et al., 2001]. More than one member could have been detected by qPCR only if the primers yielded a product of identical size from more than one transcript, which is possible given 15 or more copies of LCR16a in the human genome.

LCR16a is thought to have been duplicated very recently in evolution and evolved rapidly, suggesting there may not be orthologues in other species. I corroborated this by a BLAST search of the most recent dbEST (July 2004) using AF132984. No significant hits ($E<0.1$) were found in non-primate species. *Morpheus* is highly represented in several libraries with more than seven hundred ESTs. NPIP shows over 90% identity with regions of chimpanzee, 60% or more identity with regions of Kloss's gibbon, orangutan and vervet monkey NPIP-like proteins and similarity with a number of other Old World monkey NPIP-like proteins (HomoloGene, BLAST). Apart from about 25% similarity with chicken caldesmon, a drosophila predicted protein and a portion of an archeal coiled-coil domain, no significant similarities with any other known protein were revealed by BLAST, PROSITE or motif-based domain searching software. Johnson et al. determined eight exons for NPIP by genomic sequence analysis. They also reported a single putative trans-membrane domain surrounded by alpha-coil in AAD34394. I detected a putative transmembrane domain from amino acid 54-72 and also a putative coiled-coil domain from amino acid 139-172 using several prediction programs.

I raised two anti-sera to the C-terminus of NPIP. It is likely that antibodies raised this way (even if they were monoclonal) will cross-react with other members of the *morpheus*

family due to sequence similarity. However, this will be desirable until the members that are expressed in retina are found. It is also possible that only one member is not a pseudo-gene. Initial characterization was difficult as there are no reference cells or tissues known to express *morpheus* proteins in large quantities.

I cloned and sequenced several cDNAs from both brain and retina amplified using primers (Table 2.3) to yield full-length NPIP. Note that the forward primers both begin at the START codon as this was the most 5' high-quality sequence found. My sequence data (Appendix B Figure B.6, Figure B.5) provide evidence for at least two novel isoforms of NPIP.

The first novel isoform (NPIP-brain1) I had high-quality bi-directional sequence for differs from the reference AF132988 by an extra 47 nucleotides. This extra sequence is within exon 8 if the genomic structure determined by Johnson et al. within AC002045 is used. However, it is not known exactly where in chromosome 16 real transcripts are derived from. Although a 47 nucleotide insertion would be expected to produce a frame-shift, the resulting translation is a protein that has an identical C-half but is 19 amino acids longer. This illustrates the repetitive nature of exon 8. I found both NPIP-brain1 and AF132988 match the human genome sequence in several locations in chromosome 16p using BLAST, BLAT and by manually aligning resultant sequence. Neither appear to match 100% of bases at any location in their entirety as well as fitting into orthodox AT-GT splice sites for each putative exon. However, I also found a large number of ESTs containing the 'extra' sequence.

The second novel isoform (NPIP-brain2), that I obtained high-quality bi-directional sequence for contains a longer exon 1 that I refer to as exon 1A. It fits in with the exon structure of NPIP derived from AC002045 by using an orthodox alternative splice site. However, there are several differences in exon 6. A translation of reading frame 2 (relative to AF132984 and NPIP-brain1) results in the same protein as NPIP-brain1 minus the first 123 amino acids and with 4 amino acid changes due to exon 6. I obtained high-quality sequence from another clone that also contained exon 1A as well as several corresponding

EST hits. However, I did not find any other evidence for the expression of the 4 amino acid changes.

It is clear that AF132984 and my novel transcripts may potentially have originated from one or more members of the *morpheus* gene family in chromosome 16. Difficulty in pinning down transcripts to genomic locations may well reflect problems with the genome sequence due to micro and macro repeats, problems with cloning and sequencing transcripts from human tissue, and genuine polymorphism. However, even if a large number of *morpheus* clones were sequenced it may still be difficult to determine their genomic origin by this method or by rapid sequencing of cDNA ends (RACE). A single species of *morpheus* transcript may also be derived from more than one genomic location. Studying the promoter regions may help in exclusion of pseudo-genes and in determining tissue specific expression. Linkage studies and mutational analyses may reveal which *morpheus* gene is expressed in human macula.

Sub-cloning of NPIP-brain1 into pGEX as a GST fusion protein for prokaryotic expression allowed me to optimize conditions for Western blotting. Subsequently, I showed expression of *morpheus* in both peripheral and whole retina Figure 6.1). I detected bands suggestive that both a long ≈40 kDa and short ≈25 kDa isoform are expressed in human retina. The larger band corresponds to the predicted size for NPIP-brain1 and the smaller band for NPIP-brain2. However, the smaller band could have been due to degradation and this could also explain why the signal in retina was not overwhelming. The antibody was not perfect for Western blotting as it was difficult to detect a specific signal in cultured cells, although they had been boiled and freeze-thawed several times so degradation could be a factor, especially if the protein was susceptible to protease. Interestingly, I found significant PEST sequences (PEST score >5) in AAD34394, NPIP-brain1 and NPIP-brain2. PEST polypeptides – proteins rich in proline (P), acidic residues glutamic acid (E) or aspartic acid, and hydrophilic residues serine (S) and threonine (T) are degraded rapidly inside eukaryotic cells [Rogers et al., 1986] [4]. This could be one reason that there is such

[4]Duplicated PEST domains have been found in some proteins [Chevaillier, 1993] and are an example of regional rather than gene duplication.

a large amount of *morpheus* mRNA in the retina, especially in the macula which has the highest turnover.

I showed that the anti$_b$-NPIP antibody was useful for immunofluorescence studies by staining endogenous protein in un-transfected human neuroblastoma cells (Figure 6.2). It showed a very similar pattern of staining around the nuclear pore to that of a GFP construct containing AF132984 [Johnson et al., 2001, supplementary data]. It was a little less specific, although that is partly expected from a polyclonal antibody. It is also possible that members of the *morpheus* family not localized to the nuclear pore are detected by the antibody.

I found different localization of my 2 novel *morpheus* isoforms using GFP expression constructs. NPIP-brain2 definitely did not localize to the nuclear pore, while NPIP-brain1 was present in the region of the nuclear envelope. However, SK-N-SH cells tend to round up when transfected with some constructs, which squashes the cytoplasm towards the nucleus making it very difficult to definitively distinguish the nuclear envelope, hence the images are not shown. It is likely that expression in another cell line, such as the COS cells used by Johnson et al. would yield clearer nuclear pore localization of NPIP-brain1. NPIP-brain2 is missing a large part of the N-terminus compared with AAD34394 and NPIP-brain1, in which there is a single putative transmembrane domain. This provides further evidence that AAD34394 and NPIP-brain1 are membrane-bound proteins. There may also be a nuclear pore localization signal in the N-terminus.

There was very specific localization of *morpheus* in the human retina. In initial studies of macular sections a sub-population of photoreceptors were labelled by my antibody that looked more like cones than rods. Use of a superior secondary antibody and co-localization with PNA in the human fovea revealed that *morpheus* is expressed in cone outer segments (Figure 6.4). My imaging studies and evidence that at least one member of the gene family is a membrane-bound protein infer specific localization of *morpheus* to the cone outer segment membrane. This suggests that it may have a function related to the photo-transduction cascade or its maintenance.

Mutations in the *morpheus* gene family may cause monogenic macular, cone, or cone-rod dystrophy. Alterations in the sequence or expressed 'dose' of one or more family members may also contribute to the pathogenesis of age-related macular degeneration. Autosomal dominant retinitis pigmentosa (adRP) has been linked to a chromosomal region that overlaps with *morpheus* (16p12.3-p12.1) [Finckh et al., 1998]. However, one of the two families showed "physical signs frequently seen in patients with (Bardet-Biedl syndrome) BBS". The other family reported night-blindness from the second decade, which is characteristic of rod-cone dystrophy, although no retinal electrophysiology was published. At least two members of the *morpheus* gene family are located within a region of suggested linkage (16p12) obtained from three separate genome-wide scans of age-related maculopathy [Schick et al., 2003; Iyengar et al., 2004; Schmidt et al., 2004].

This is the first report, to my knowledge, of tissue-specific localization of the *morpheus* gene family. It is interesting that it seems to be limited to cone photoreceptors. Although longer isoforms of NPIP appear to be expressed close to the nuclear pore, it is possible that this member of the *morpheus* family has an alternative expression pattern in photoreceptors. *Morpheus* certainly appears to be highly expressed at an mRNA level in the retina (chapter 4), especially in the macula. There is evidence that RanBP2 – a large protein that binds the nuclear pore associated GTPase Ran – is highly expressed in cone photoreceptors of the vertebrate retina [Ferreira et al., 1996]. Indeed, it is thought to be a chaperone for red/green cone opsin.

6.4 Conclusions

I was able to produce an antibody to NPIP that was useful and specific in both Western blotting and immunohistochemistry. By cloning NPIP from retinal cDNA I found two novel transcripts that are isoforms or different members of the *morpheus* gene family. Additionally, I expressed these novel transcripts and found that they localize to different parts of the cell. I detected both proteins in human retina by Western blotting. I also showed very specific expression of *morpheus* in cone outer segments by immunofluores-

cence. These observations suggest a novel role for *morpheus* in the macula. It may be important in relation to phototransduction due to its location. This is the first implication of *morpheus* playing a part in retina-specific biology. In summary, I have made progress towards eliciting the function of *morpheus* in the macula and demonstrated that it is a very interesting candidate for macular disease.

6.5 Overall conclusions

I procured a human macular resource that was suitable for gene expression studies and went on to rigorously combine multiple microarray experiments with quantitative PCR analysis. I identified several genes enriched in the macula and made progress in characterizing the proteins encoded by histone deacetylase 9 and the *morpheus* gene family, both of which are expressed in cone photoreceptors. Several other genes also provided insight into the mechanisms of precise vision and its maintenance.

I met the overall objectives to find genes that are differentially expressed in the macula and to characterize one or more of the proteins encoded by these genes. Several genes I identified by this approach are novel in the context of macular expression and are excellent candidates for macular disease.

6.6 Future work

There is a great deal of future work that could stem from this thesis. There are many potential candidate genes for macular disease to study and many genes that could lead to insight into the function of the human macula. One difficulty is knowing which of the many genes implicated are worthy of further study.

6.6.1 Gene expression

In an ideal world with exhaustive amounts of fresh human retinal tissue and other resources, it would be very useful to repeat both the microarray and quantitative PCR ex-

periments on a much larger scale. A detailed expression map of the retina, with accurate expression levels in average individuals of specific age groups would be an excellent foundation for studying the molecular physiology and pathology of the retina.

Another logical step is to study both maculas of different age and the differences between gene expression in AMD and non-AMD eyes. Perhaps more genes could also be found for study amongst the many that were found to be differentially expressed according to my microarray data, by applying less stringent exclusion criteria. However as others have also shown, microarray data is inherently noisy and there is a relatively large amount of variation across different samples. [Ross et al., 2000] It would be best to combine my microarray data with independent data from another group or further set of experiments.

6.6.2 Mutation screening

It would be very useful to look for mutations in genes I have identified as candidates for macular disease in panels of DNA from patients who suffer from these diseases. Work is currently underway by Mark Lathrop's group at the Centre Nationale Genotypage in France to study polymorphism in HDAC9, GCN2 and X in a panel of over 1500 AMD patients and many controls. The *morpheus* gene family is too large to study in its entirety so further work is needed to accurately determine which members are of most importance in the macula.

To the best of my knowledge, there are currently no loci for monogenic macular disease which correspond to any of the genes I confirmed to be differentially expressed in central retina for which a panel is available. One locus at 16p12.3-1 RP22 [Finckh et al., 1998] does correspond with the *morpheus* locus but the panel was not available anymore. Also, the phenotype was more peripheral than central retinal and it is likely that one of the two families carried a Bardet-Biedl mutation at 16q21 (OMIM #209900).

One further approach would be to screen panels of unrelated patients with cone, cone-rod, or macular dystrophies, in which a Mendelian inheritance pattern is likely, by association studies – for example by sequencing exons. The chance of finding a polymorphism

that is associated with an unknown dystrophy is small for any one gene and would require many panel members and controls to be studied. This approach may be more effective if many genes could be screened for at once – for example by using an array based re-sequencing approach. Panels of patients with cone, cone-rod and macular dystrophies are being assembled at Moorfields and The Institute of Ophthalmology.

6.6.3 Functional models

One direction of study that would be extremely useful in determining the physiology and pathology of the macula is to dissect the function of the macular genes I identified. Although there is no primate cell culture model of cone photoreceptors that I know about, some macular genes I identified may be applicable to study in avian photoreceptor cultures (chickens do not have a macula although some other bird species have either single or dual maculas).

Animal models may also be useful *in vitro*. For example, an HDAC9 knockout mouse may exhibit retinal pathology – although mice do not have maculas. Knockout mice would also not be applicable to the *morpheus* gene as it is primate specific. Other animal models may also be useful for some genes, such as the tetraploid knock-in *Xenopus*.

A particular function may be elicited at the cellular level for one or more macular genes, either by animal models, disease mutations or by direct assays that may be applicable to genes further down the list on which I have not completed protein work. If a particular function is suspected, RNA inhibition may be a useful tool to show evidence for such a role.

Acknowledgements

Supervisors

I would like to thank Alan Bird, my main supervisor for providing inspiration to me in carrying out my research and for being there for advice when I needed it. I am extremely grateful to my secondary supervisor Andrew Webster for his day-in day-out commitment to furthering my research. Andrew has been a great inspiration to me and a constant fountain of knowledge and ideas. I would also like to thank my head of department, Shomi Bhattacharya, for providing an excellent environment and for his input to my research.

I am very grateful to my co-supervisors, who put in so much of their time and effort in providing continuous help and support to my research. I would especially like to thank Bob Molday, who looked after me during my time in Canada, and Alison Hardcastle and Mike Cheetham, who looked after me in The Institute of Ophthalmology. In addition, I am grateful to Russell Foster for enabling me to carry out quantitative PCR experiments in his laboratory.

Colleagues

I would like to thank all the post doctoral scientists, laboratory technicians, graduate and undergraduate students I worked with in the many labs I was involved in both in Canada and the UK. They are too numerous to mention all of them here individually, but I would especially like to thank Laurie Molday and Chris Loewen in Bob Molday's lab, Lili Chen in Andrew Webster's lab, Suba Poopalasundaram, Neil Ebeneezer and Simon Brooks in

Alison Hardcastle's lab, Jacqui Van der Spuy and Paul Chapple in Mike Cheetham's lab and Stuart Peirson in Russell Foster's lab.

Friends and family

I am greatly indebted to all my friends and family for being so understanding and supportive over the last few years. I would especially like to thank my Mum for being such a loving personal manager and administrator and my Dad for caring so much about my future and providing perspective on what is important in life.

I would also like to thank St John's College at the University of British Columbia for making my stay in Vancouver so enjoyable and Ed Keohane, Mike Ashby and Liz van Holten my close friends and flatmates in London for putting up with me and providing ongoing intellectual, psychological and social input.

Appendix A

Scripts

A.1 Microarray data calibration

```
# source("analysis.r")
library(sma)
library(vsn)
source("background_correction.r")

barcode <- "016.txt/16"

# cy3:
# read in data
input.file <- paste(barcode,"_1.txt",sep="")
cy3 <- background.correct(input.file)

# cy5:
# read in data:
input.file <- paste(barcode,"_2.txt",sep="")
cy5 <- background.correct(input.file)

layout <- NULL
layout$ngrid.r <- 12
layout$ngrid.c <- 4
```

```
layout$nspot.r <- 23
layout$nspot.c <- 26

slide <- NULL
slide$G <- cbind(cy3$Signal.Median-cy3$Background.Adjustment)
slide$Gb <- NULL
slide$R <- cbind(cy5$Signal.Median-cy5$Background.Adjustment)
slide$Rb <- NULL

slide.vsn <- cbind(slide$R, slide$G)

# Spots below background will not be counted in the analysis
# Include these 2 lines if you want to set them to 1
#slide$G[slide$G<1] <- 1
#slide$R[slide$R<1] <- 1

log.ratios <- stat.ma(slide, layout, norm="n")
log.ratios.median <- stat.ma(slide, layout, norm="m")
log.ratios.lowess <- stat.ma(slide, layout, norm="l")
log.ratios.tip <- stat.ma(slide, layout)
log.ratios.scale <- stat.ma(slide, layout, norm="s")
res <- vsn(slide.vsn)

par(mfrow=c(2,1))
plot(log.ratios$A, log.ratios$M)
plot(log.ratios.median$A, log.ratios.median$M)
plot(log.ratios.lowess$A, log.ratios.lowess$M)
plot(log.ratios.tip$A, log.ratios.tip$M)
plot(log.ratios.scale$A, log.ratios.scale$M)
plot((res$hy[,1]+res$hy[,2])/2, res$hy[,1]-res$hy[,2])

columns <- c(1,2)
cols <- c("red", "blue")[as.numeric(res$sel)+1]
cnam <- c("Cy5","Cy3")
```

```
rtxty <- rank(res$havg)
dtxty <- res$hy[,2]- res$hy[,1]

plot(rtxty, dtxty, pch=".", col=cols,
ylim=quantile(dtxty, probs=c(1e-3, 1-1e-3)),
xlab="rank␣of␣mean␣expression", ylab=paste("difference", cnam[2], "-",
    cnam[1]))
lines(1e6*c(-1,1), c(0,0), col="black", lty="dashed")

logrank <- rank(log.ratios.median$A)
plot(logrank, -log.ratios.median$M, pch=".", col="blue",
xlab="rank␣of␣mean␣expression", ylab=paste("difference", cnam[2], "-",
    cnam[1]))
lines(1e6*c(-1,1), c(0,0), col="black", lty="dashed")

array1 <- data.frame(cy3$Index,cy3$Accession,cy3$Text,cy3$Signal.Median
    ,cy3$Background.Adjustment,cy5$Signal.Median,cy5$Background.
    Adjustment,log.ratios$A,log.ratios$M,log.ratios.median$M,log.ratios
    .lowess$M,log.ratios.tip$M,log.ratios.scale$M,res$hy[,2],res$hy
    [,1],res$hy[,2]-res$hy[,1])

names(array1) <- c("id","accession","name","cy3","cy3.bg","cy5","cy5.bg
    ","expression","log.ratio","lr.median","lr.lowess","lr.tip","lr.
    scale","cy3.vsn","cy5.vsn","vsn.ratio")

save(list=ls(),file=paste(barcode,".rdata",sep=""))

# write for spreadsheet
write.table(array1,"012.txt/slide12.csv",col.names=TRUE,quote=FALSE,sep
    ="\t",row.names=FALSE)

array1$expression[!is.finite(array1$expression)] <- "NULL"
array1$log.ratio[!is.finite(array1$log.ratio)] <- "NULL"
```

```
array1$lr.median[!is.finite(array1$lr.median)] <- "NULL"
array1$lr.lowess[!is.finite(array1$lr.lowess)] <- "NULL"
array1$lr.tip[!is.finite(array1$lr.tip)] <- "NULL"
array1$lr.scale[!is.finite(array1$lr.scale)] <- "NULL"

# write for SQL
write.table(array1,"012.txt/slide12",col.names=FALSE,quote=FALSE,sep="\
    t",row.names=FALSE)
```

A.2 Microarray data mining

A.2.1 Final table design

```
DROP DATABASE array
CREATE DATABASE array;
USE array;

DROP TABLE slide_n;
CREATE TABLE slide_n (
    id INT UNSIGNED NOT NULL PRIMARY KEY,
    accession VARCHAR(15) NOT NULL,
    name VARCHAR(255),
    cy3 FLOAT NOT NULL,
    cy3_bg FLOAT NOT NULL,
    cy5 FLOAT NOT NULL,
    cy5_bg FLOAT NOT NULL,
    expression FLOAT NOT NULL,
    log_ratio FLOAT NOT NULL,
    lr_median FLOAT NOT NULL,
    lr_lowess FLOAT NOT NULL,
    lr_tip FLOAT NOT NULL,
    lr_scale FLOAT NOT NULL,
    cy3_vsn FLOAT NOT NULL,
    cy5_vsn FLOAT NOT NULL,
```

```
    vsn_ratio FLOAT NOT NULL,
    INDEX (name, accession)
            );
```

A.2.2 Example queries for robust over- and under-expression

```
# increased expression in fovea
select combined_15.name,
combined_15.accession,
(combined_15.cy3-combined_15.cy3_bg+combined_16.cy3-combined_16.cy3_bg)
    /2 as cy3,
(combined_15.cy5-combined_15.cy5_bg+combined_16.cy5-combined_16.cy5_bg)
    /2 as cy5,
(combined_15.vsn_ratio+combined_15.vsn_ratio_2+combined_16.vsn_ratio+
    combined_16.vsn_ratio_2)/4 as mean_vsnr_15_16
from combined_15,combined_16
where combined_15.name=combined_16.name
and combined_15.accession=combined_16.accession
and (combined_15.vsn_ratio+combined_15.vsn_ratio_2+combined_16.
    vsn_ratio+combined_16.vsn_ratio_2)/4 >= 3
and (combined_15.cy5-combined_15.cy5_bg+combined_16.cy5-combined_16.
    cy5_bg)/2 >= 200
order by mean_vsnr_15_16;

# increased expression in peri
select combined_15.name,
combined_15.accession,
(combined_15.cy3-combined_15.cy3_bg+combined_16.cy3-combined_16.cy3_bg)
    /2 as cy3,
(combined_15.cy5-combined_15.cy5_bg+combined_16.cy5-combined_16.cy5_bg)
    /2 as cy5,
(combined_15.vsn_ratio+combined_15.vsn_ratio_2+combined_16.vsn_ratio+
    combined_16.vsn_ratio_2)/4 as vsnr
from combined_15,combined_16
where combined_15.name=combined_16.name
```

```
and combined_15.accession=combined_16.accession
and (combined_15.vsn_ratio+combined_15.vsn_ratio_2+combined_16.
    vsn_ratio+combined_16.vsn_ratio_2)/4 <= -1
and (combined_15.cy5-combined_15.cy5_bg+combined_16.cy5-combined_16.
    cy5_bg)/2 >= 200
order by vsnr;
```

A.2.3 Importing data into a spreadsheet

```
#!/bin/bash
#retrieve all data for accession number=$1 (1st command-line parameter)
    to a separate file for each slide
mysql --execute="select name,accession,cy3-cy3_bg as cy3,cy5-cy5_bg as
    cy5,cy5_vsn as cy_3,cy3_vsn as cy_5,vsn_ratio from slide_11 where
    accession='$1'" array > slide11
mysql -N --execute="select name,accession,cy3-cy3_bg as cy3,cy5-cy5_bg
    as cy5,cy5_vsn as cy_3,cy3_vsn as cy_5,vsn_ratio from slide_12
    where accession='$1'" array > slide12
mysql -N --execute="select name,accession,cy3-cy3_bg as cy3,cy5-cy5_bg
    as cy5,cy5_vsn as cy_3,cy3_vsn as cy_5,vsn_ratio from slide_13
    where accession='$1'" array > slide13
mysql -N --execute="select name,accession,cy3-cy3_bg as cy3,cy5-cy5_bg
    as cy5,cy5_vsn as cy_3,cy3_vsn as cy_5,vsn_ratio from slide_14
    where accession='$1'" array > slide14
mysql -N --execute="select name,accession,cy3-cy3_bg as cy3,cy5-cy5_bg
    as cy5,cy5_vsn as cy_3,cy3_vsn as cy_5,vsn_ratio from slide_15
    where accession='$1'" array > slide15
mysql -N --execute="select name,accession,cy3-cy3_bg as cy3,cy5-cy5_bg
    as cy5,cy5_vsn as cy_3,cy3_vsn as cy_5,vsn_ratio from slide_16
    where accession='$1'" array > slide16

#concatenate files
cat slide11 slide12 slide13 slide14 slide15 slide16 > $1.csv

#clean up
```

```
rm slide11
rm slide12
rm slide13
rm slide14
rm slide15
rm slide16
```

A.3 Graph to simulate ideal and real PCR

```
#output to encapsulated postscript file
postscript("/home/Dan/ps/qPCR.eps", horizontal = FALSE, onefile = FALSE
    )

cycle <- 1:40
#real data is sigmoidal
fluorescence <- 1/(1+2.7183^(-0.5*(cycle-30)))
plot(cycle, fluorescence, cex.axis=1.2, col='green', pch=16)
s <- seq(length(cycle)-1)
segments(cycle[s], fluorescence[s], cycle[s+1], fluorescence[s+1], col='
    green')
lines(c(25,25,0), c(0, fluorescence[25], fluorescence[25]), col='red', lty
    =2)
lines(c(29,29,0), c(0, fluorescence[29], fluorescence[29]), col='red', lty
    =2)
text(27,0.05, expression(Delta*c), col='blue', cex=1.75)
text(10,0.225, expression(Delta*f), col='blue', cex=1.75)
arrows(10,0.2,10,0.085, col='blue', length=0.1)
arrows(10,0.25,10,0.37, col='blue', length=0.1)

#theoretical amplification is exponential
f <- 100000^-1.118*1.509^s
segments(cycle[s], f[s], cycle[s+1], f[s+1], col='orange', lty=2)
legend(33.5,0.7, c('data', 'model'), col=c('green', 'orange'), lwd=1, lty=c
    (1,2), pch=c(16,256))
```

```
#turn off output
dev.off()
```

Appendix B

Supplementary data

Gene Symbol	*Description*	*Accession*	*UniGene ID*	*Tm*
ABCA4	ATP-binding cassette sub-family A (ABC1) member 4	AF000148	198396	79.4
AOC2	amine oxidase copper containing 2 (retina-specific)	AB012943	143102	77.1
ARR3	arrestin 3 retinal (X-arrestin)	S66793	308	78.8
ARRB1	arrestin beta 1	AF084040	112278	76.5
ARRB2	arrestin beta 2	AF106941	18142	80.0
BCP	blue cone pigment	M13299	102119	78.3
CALB1	calbindin 1 (28kD)	AF068862	65425	73.6
CALB2	calbindin 2 (29kD calretinin)	NM 001740	106857	81.2
CALB3	calbindin 3 (vitamin D-dependent calcium-binding protein)	L13220	639	75.3
CDH4	cadherin 4 R-cadherin (retinal)	L34059	89484	73.6
CECR1	cat eye syndrome critical region gene 1	NM 017424	170.0310	76.5
CLUL1	clusterin-like 1 (retinal)	D63813	26886	74.2
CNGA1	cyclic nucleotide gated channel alpha 1	S42457	1323	74.7
CNGA3	cyclic nucleotide gated channel alpha 3	AF065314	234785	78.3
CNGB1	cyclic nucleotide gated channel beta 1	AF042498	93909	80.0
CRX	cone-rod homeobox	AF024711	249186	80.0
FSCN2	fascin homolog 2 (actin bundling protein retinal)	AF030165	118555	80.0
GCP	green cone pigment (color blindness deutan)	M13305	247787	80.0
GUCA1A	guanylate cyclase activator 1A (retina)	L36861	92858	77.7
GUCA1B	guanylate cyclase activator 1B (retina)	M97496	778	81.2
GUCY2D	guanylate cyclase 2D membrane (retina-specific)	M92432	1974	82.4
GUCY2F	guanylate cyclase 2F retinal	L37378	123074	77.7
HCR	HCR (a-helix coiled-coil rod homologue)	AK000204	110746	78.8
IMPG1	interphotoreceptor matrix proteoglycan 1	NM 001563	129882	73.6
LOC51170.0	retinal short-chain dehydrogenase/reductase retSDR2	NM 016245	12150	74.2
LOC51171	retinal short-chain dehydrogenase/reductase retSDR3	NM 016246	18788	78.8
LRAT	lecithin retinol acyltransferase	AF071510	19470.08	73.6
None	Homo sapiens retinal mRNA	L23852	73838	70.0
None	X-linked retinopathy protein	S58722	207812	70.0
NRL	neural retina leucine zipper	M95925	89606	80.6
NTN2L	retinaldehyde-binding protein 1	U86758	158336	81.2
OCA2	oculocutaneous albinism II	NM 000275	82027	76.5
PDE2A	phosphodiesterase 2A cGMP-stimulated	U67733	154437	75.9
PDE5A	phosphodiesterase 5A cGMP-specific	D89094	139271	75.3
PDE6A	phosphodiesterase 6A cGMP-specific rod alpha	M26061	182240	75.9
PDE6B	phosphodiesterase 6B cGMP-specific rod beta (adCSNB3)	S41458	2593	76.5
PDE6C	phosphodiesterase 6C cGMP-specific cone alpha prime	U31973	93173	75.9
PDE6D	phosphodiesterase 6D cGMP-specific rod delta	AF022912	48291	75.9
PDE6G	phosphodiesterase 6G cGMP-specific rod gamma	X62025	1857	81.2
PDE6H	phosphodiesterase 6H cGMP-specific cone gamma	D45399	54471	75.9
PRRDH	photoreceptor outer segment all-trans retinol dehydrogenase	AF229845	272405	78.3
RALDH2	retinaldehyde dehydrogenase 2	AL110274	95197	75.3
RBP1	retinol-binding protein 1 cellular	NM 002899	101850	78.3
RBP2	retinol-binding protein 2 cellular	NM 004164	182313	80.0
RBP3	retinol-binding protein 3 interstitial	M33875	857	82.4
RBP4	retinol-binding protein 4 interstitial	NM 006744	76461	74.2
RCV1	recoverin	S43855	80.0539	79.4
RDGBB	retinal degeneration B beta	NM 012417	109219	76.5
RDH5	retinol dehydrogenase 5 (11-cis and 9-cis)	AF0370.062	172914	74.2
RDHL	retinol dehydrogenase homolog	NM 005771	179608	77.7
RDS	retinal degeneration slow (retinitis pigmentosa 7)	M73531	1937	79.4
RGR	retinal G protein coupled receptor	U14910	1544	75.9
RHO	rhodopsin (retinitis pigmentosa 4 autosomal dominant)	NM 000539	247565	76.5
RHOK	rhodopsin kinase	U63973	103501	81.8
RLBP1	retinaldehyde-binding protein 1	L34219	1933	76.5
RODH	retinol dehydrogenase	U89281	11958	75.9
RODH-4	microsomal NAD+-dependent retinol dehydrogenase 4	AF0570.034	134958	79.4
ROM1	retinal outer segment membrane protein 1	M96759	189852	78.8
RP1	retinitis pigmentosa 1 (autosomal dominant)	NM 006269	251687	73.6
RP2	retinitis pigmentosa 2 (X-linked recessive)	AJ007590	44766	75.3
RPE65	retinal pigment epithelium-specific protein (65kD)	U18991	2133	73.1
RPF-1	Retina-derived POU-domain factor-1	U91935	233321	78.3
RPGR	retinitis pigmentosa GTPase regulator	AJ238395	153614	74.2
RS1	retinoschisis (X-linked juvenile) 1	AF014459	113250	80.0
RX	retinal homeobox protein	NM 013435	278957	75.3
SAG	S-antigen; retina and pineal gland (arrestin)	X12453	32721	77.1
SCA7	spinocerebellar ataxia 7	AJ000517	108447	75.9
SPACRCAN	interphotoreceptor matrix proteoglycan 200	NM 016247	272380.0	78.3

Table B.1: A search for retina-related genes amongst the 13,899 probes in the microarray revealed over 50.

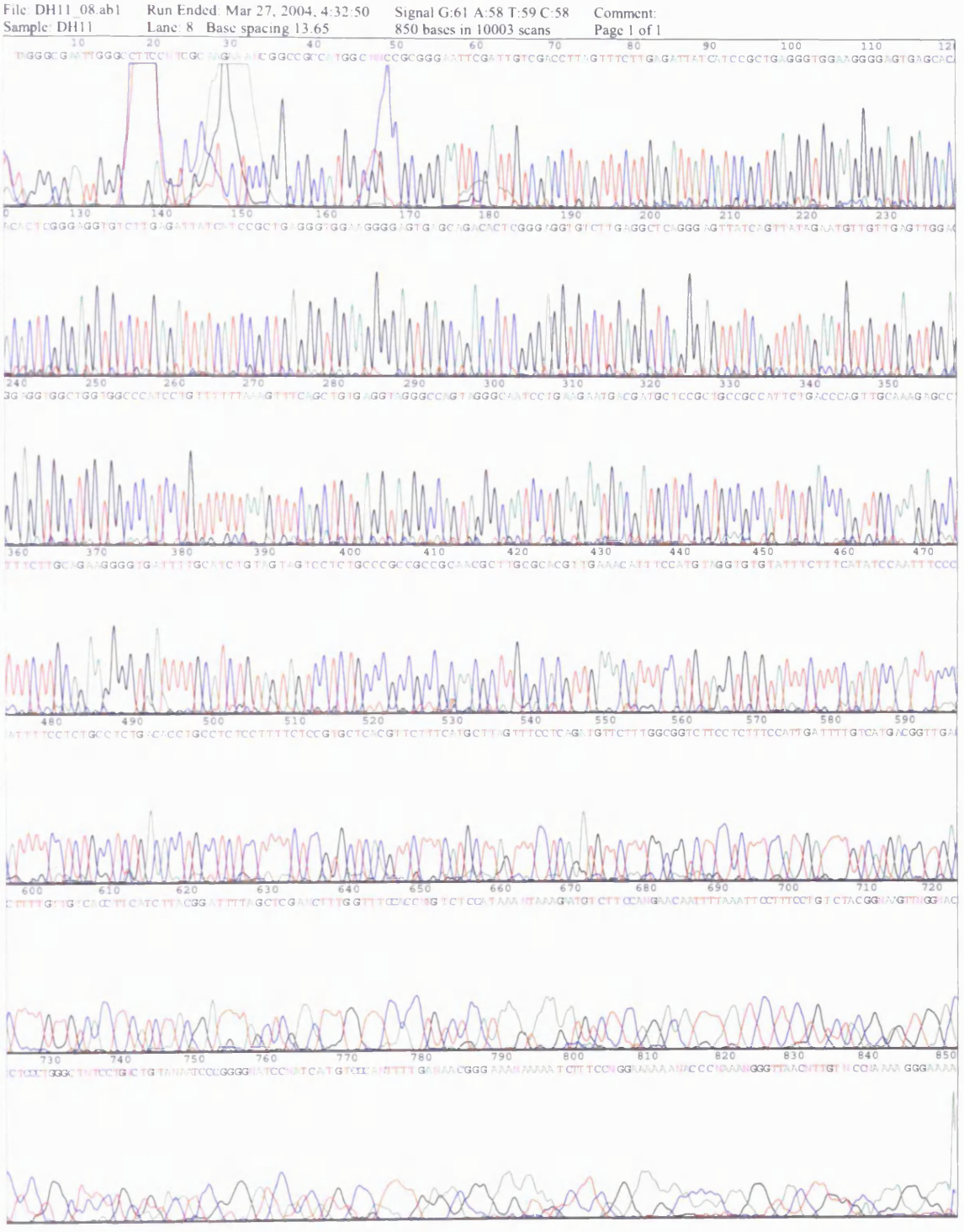


Figure B.1: Direct sequencing of npip-brain1 cDNA (forward strand).

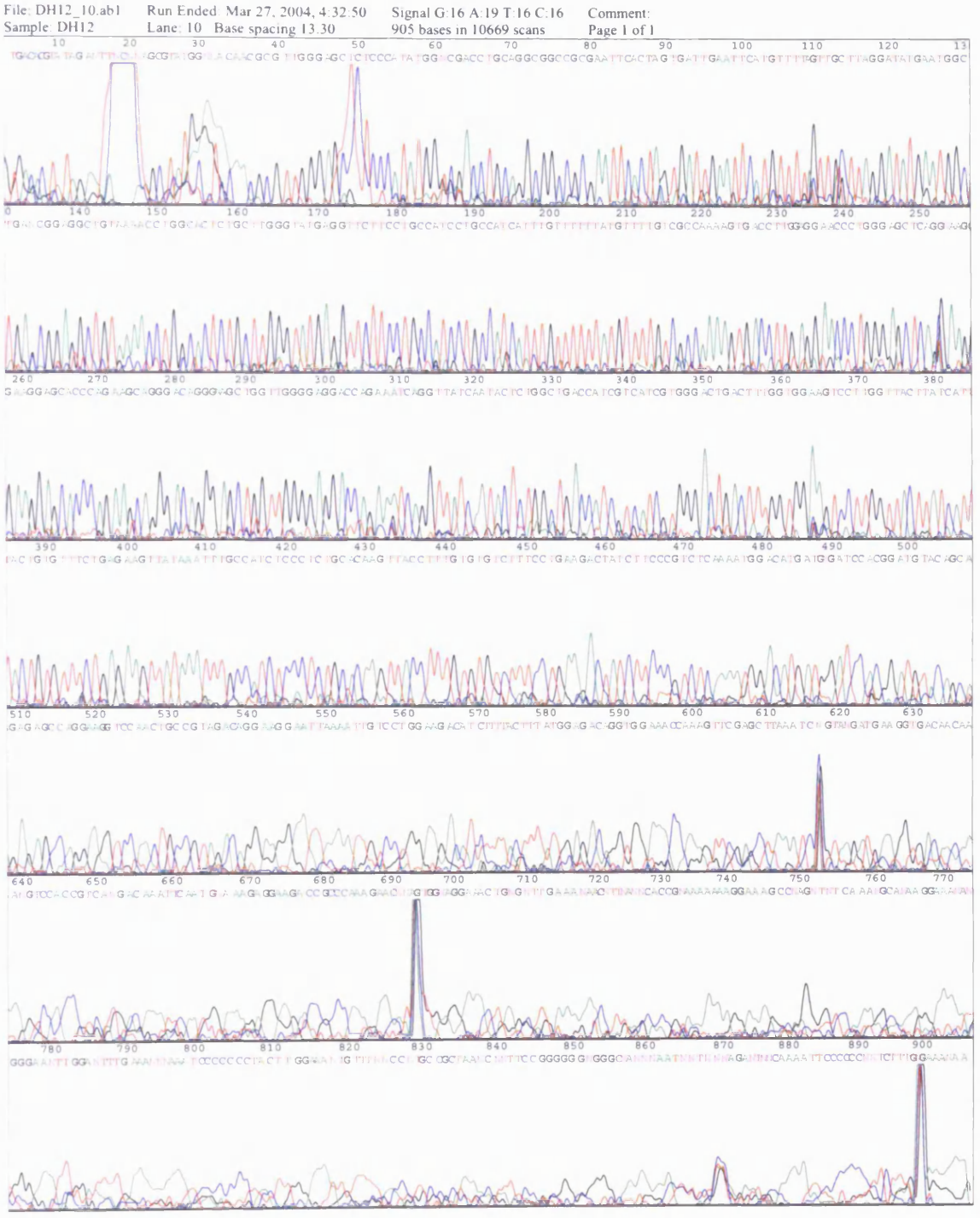


Figure B.2: Direct sequencing of npip-brain1 cDNA (reverse strand).

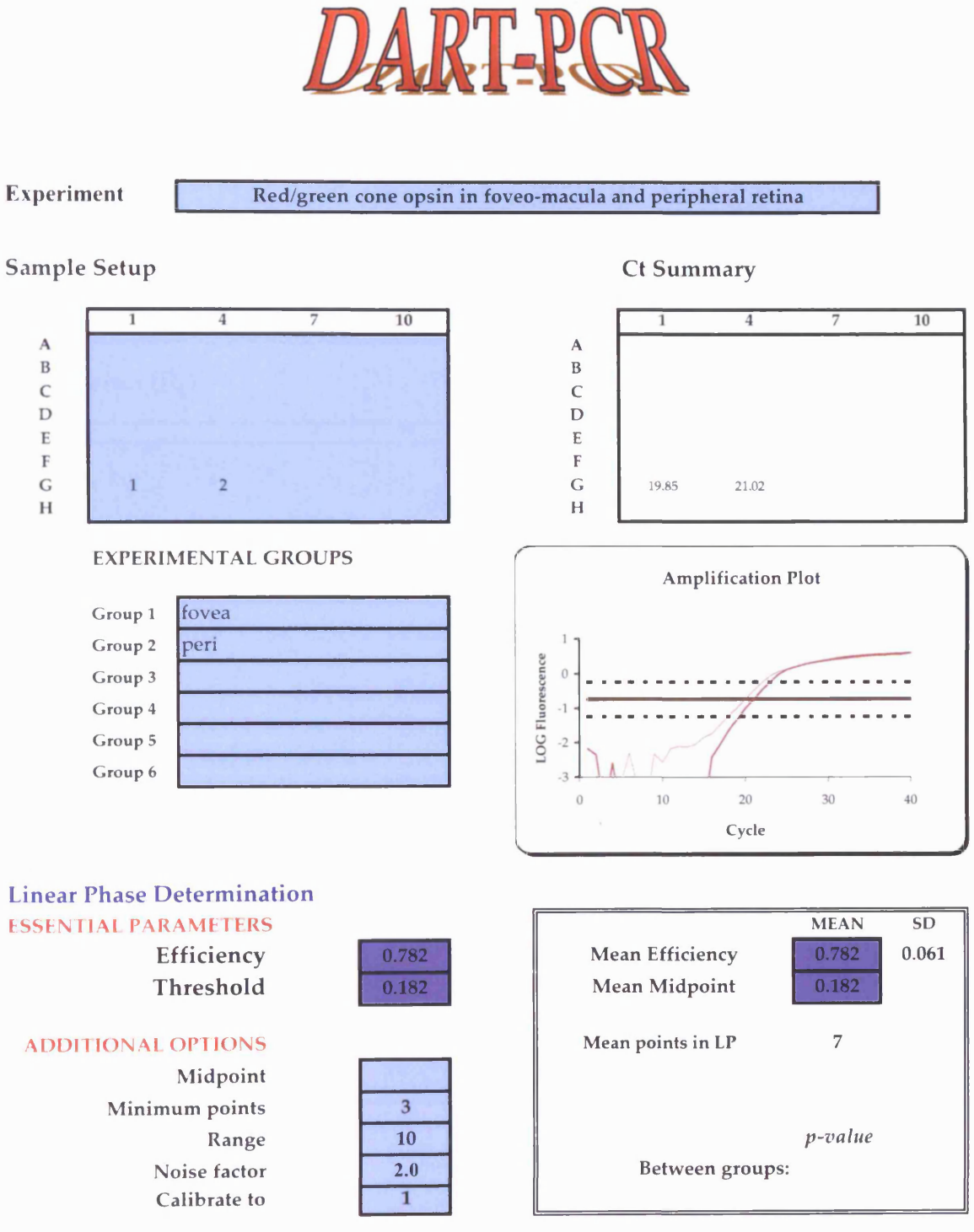
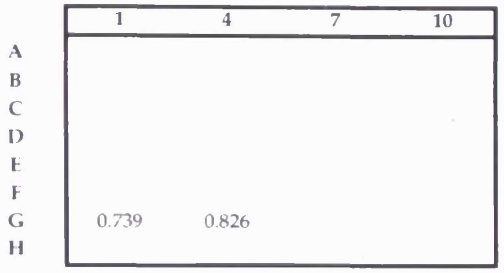


Figure B.3: DART-PCR for red/green cone opsin

Results

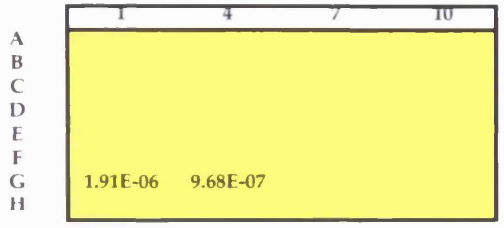
Efficiency Results



Experimental Group Efficiency

Sample	E	SD	p-value
1	0.739		
2	0.826		
3			
4			
5			
6			

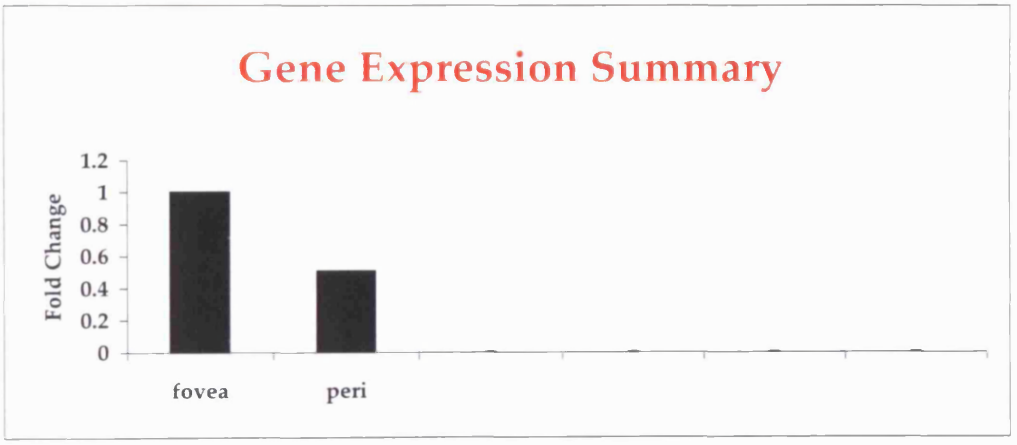
Expression (R_0)



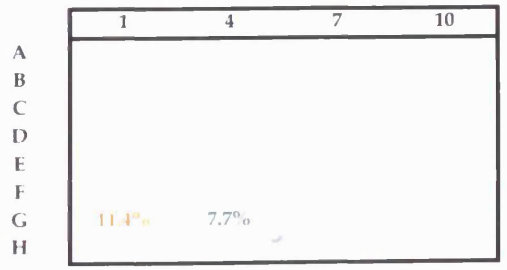
Fold Change

Group	Fold	SD
1	1.000	
2	0.505	
3		
4		
5		
6		

Gene Expression Summary

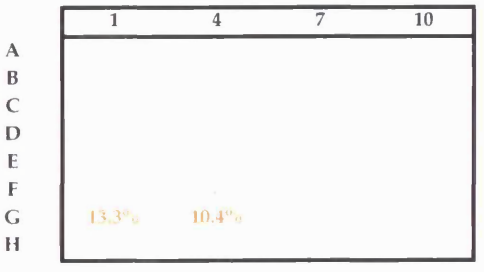


Expression CV



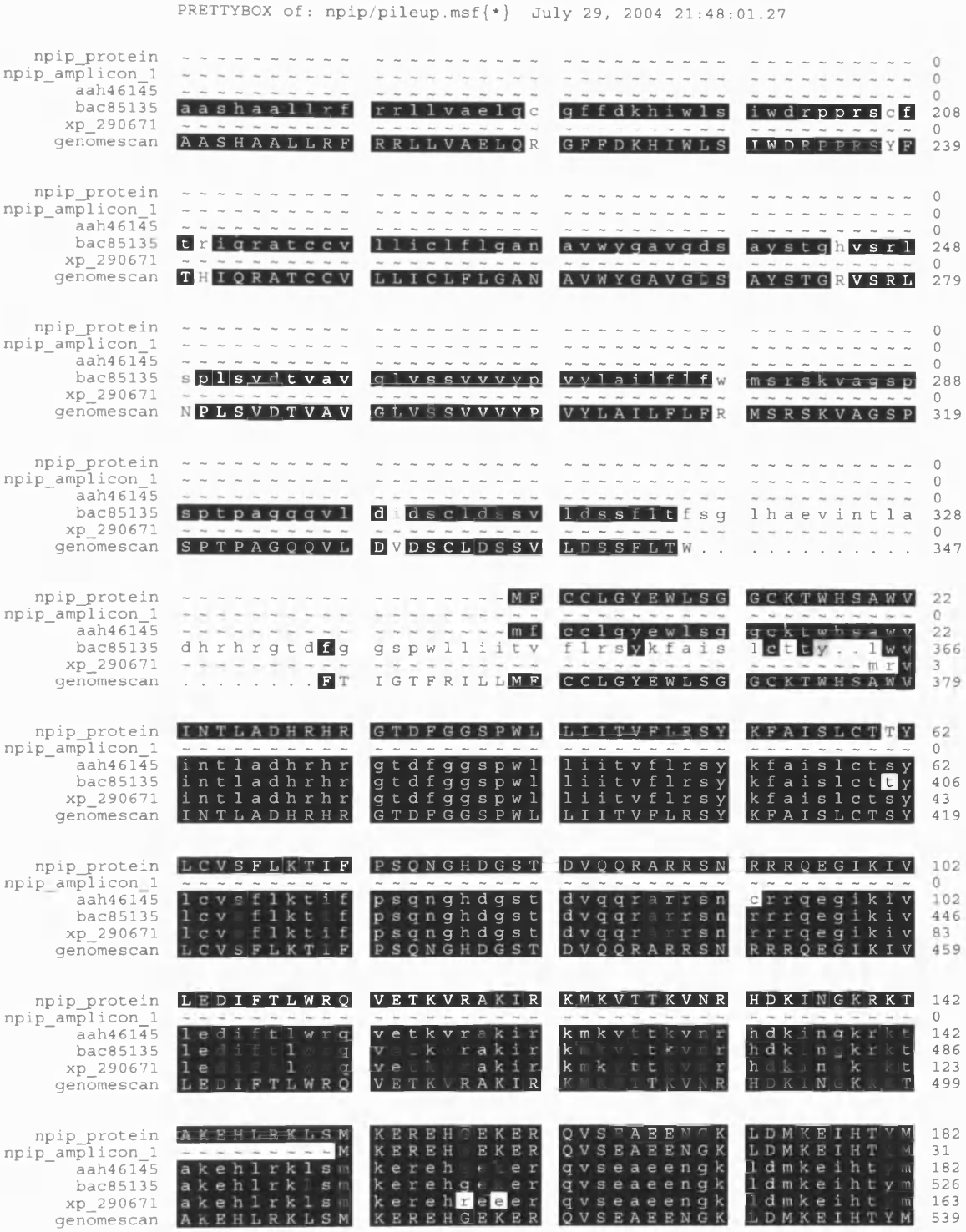
Mean CV 9.6%

Efficiency CV



Mean CV 11.8%

Figure B.4: DART-PCR for red/green cone opsin


PRETTYBOX of: npip/pileup.msf{*} July 29, 2004 21:48:01.27
npip_protein
npip_amplicon_1
aah46145
bac85135
xp_290671
genomescan

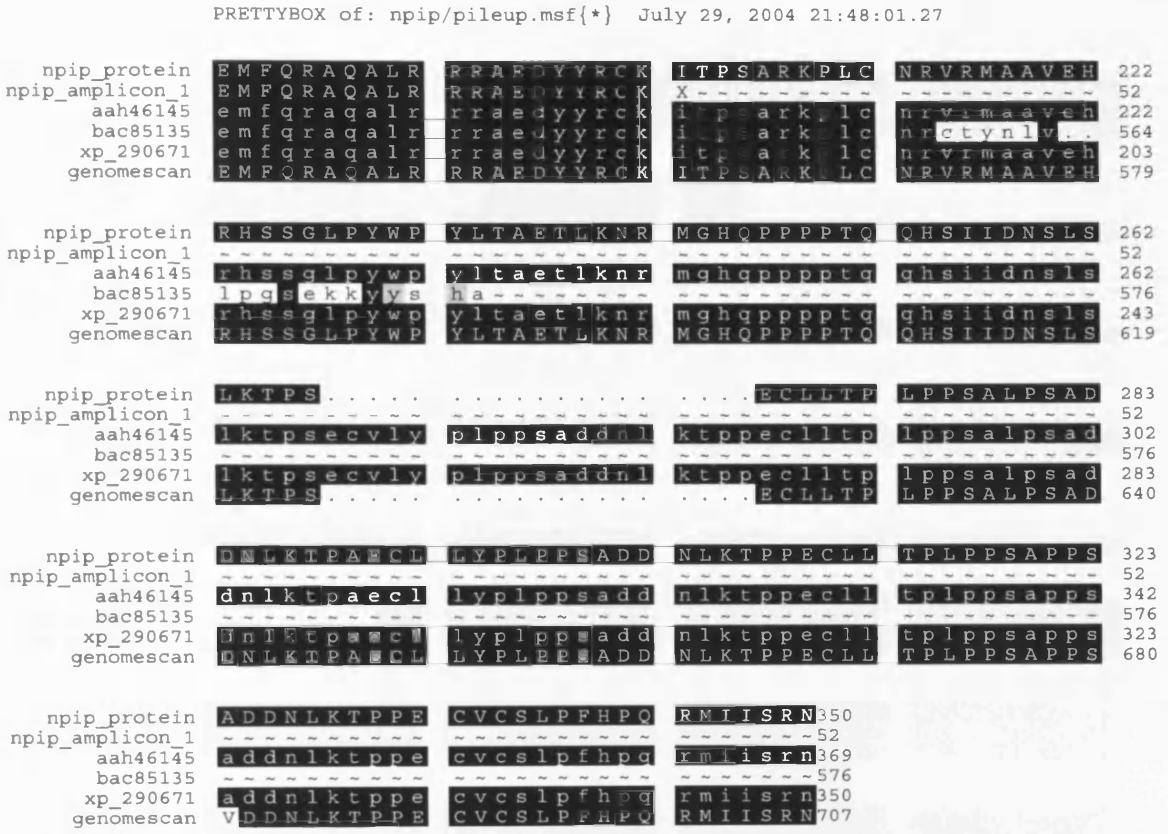
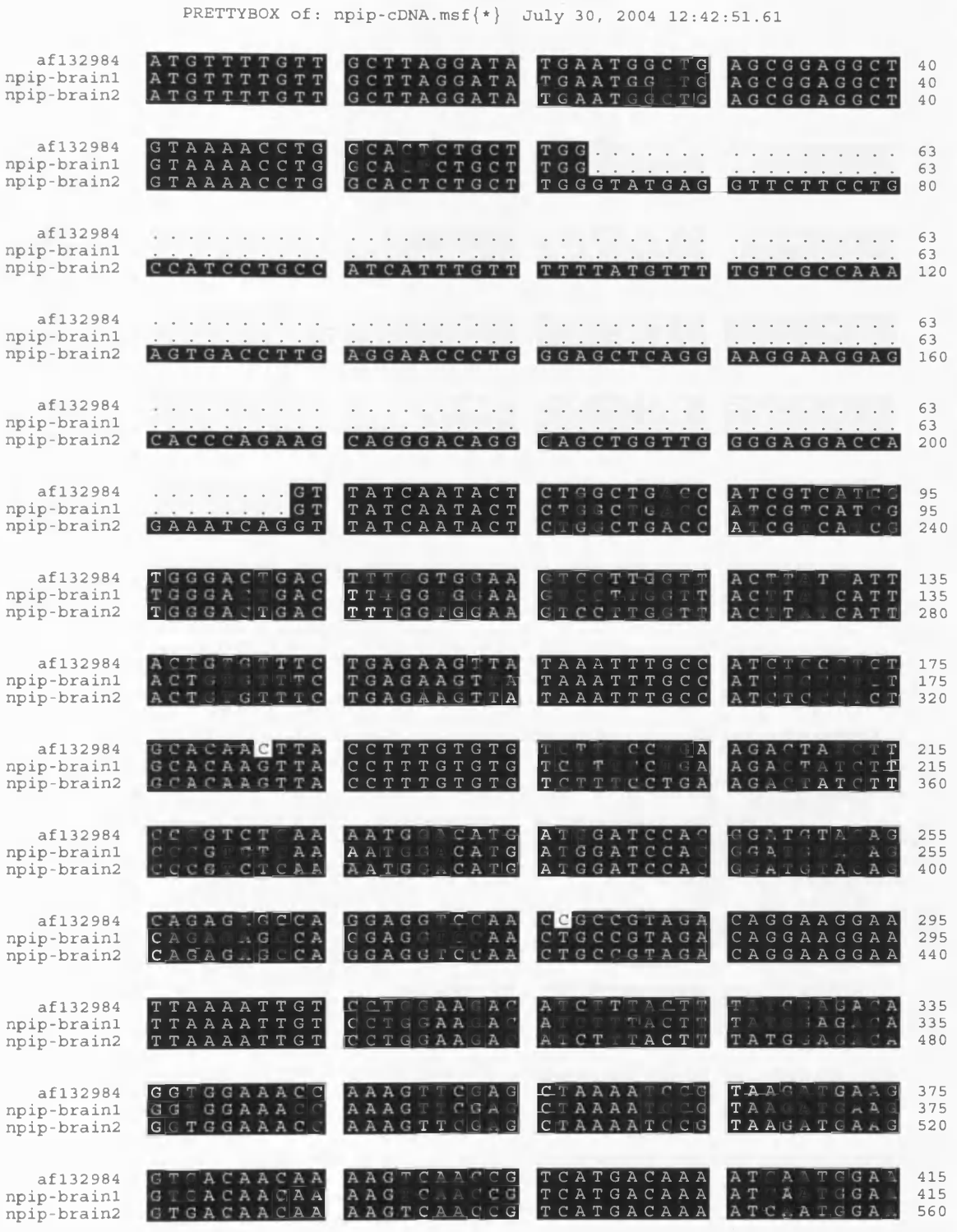
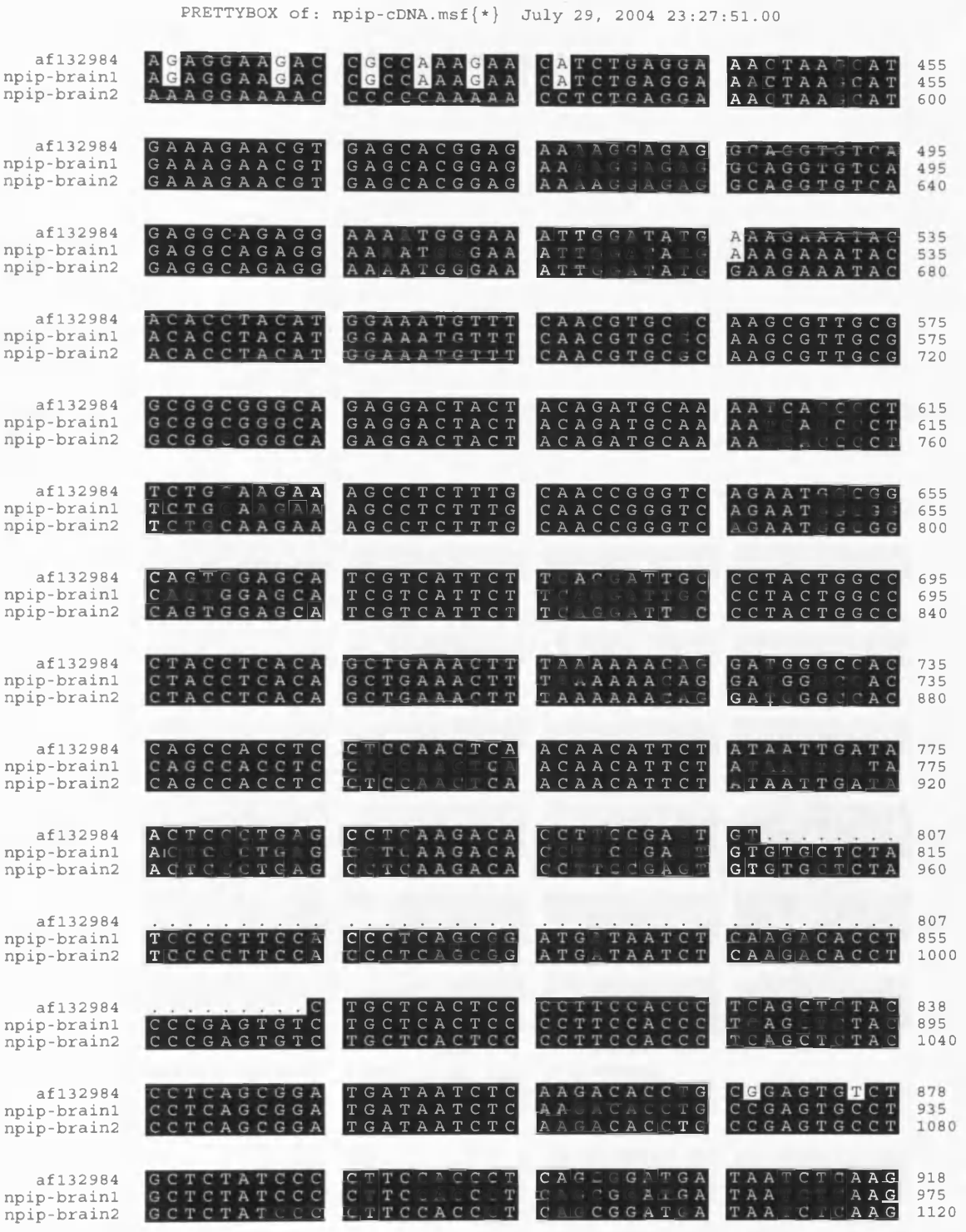


Figure B.5: Multiple alignment of NPIP protein (AAD34394) with the amplicon used for qPCR, various NPIP-like proteins, and a predicted protein from a BAC clone covering 16p13.1.


PRETTYBOX of: npip-cDNA.msf{*} July 30, 2004 12:42:51.61
af132984
npip-brain1
npip-brain2


PRETTYBOX of: npip-cDNA.msf{*} July 29, 2004 23:27:51.00
af132984
npip-brain1
npip-brain2
455
455
600
495
495
640
535
535
680
575
575
720
615
615
760
655
655
800
695
695
840
735
735
880
775
775
920
807
815
960
807
855
1000
838
895
1040
878
935
1080
918
975
1120

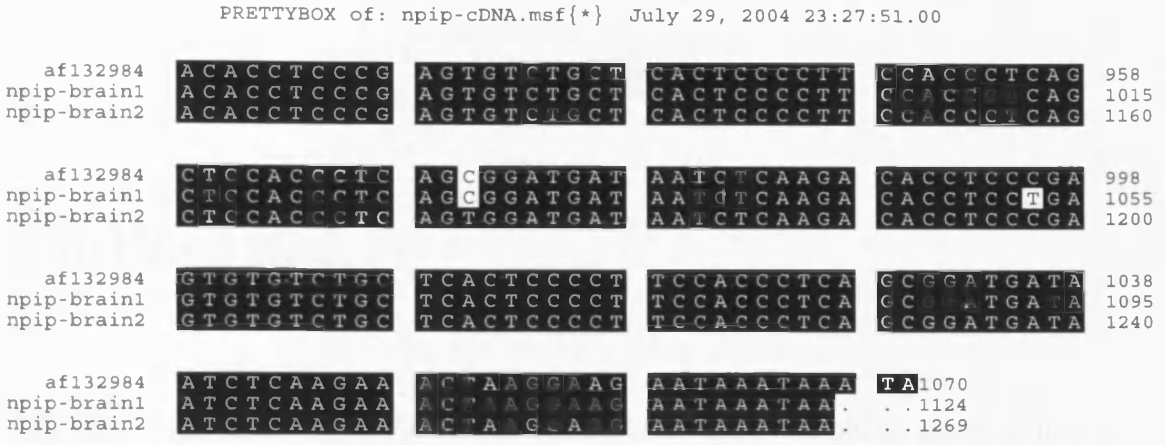


Figure B.6: Multiple alignment of NPIP reference cDNA AF132984 with 2 cDNAs found in brain.

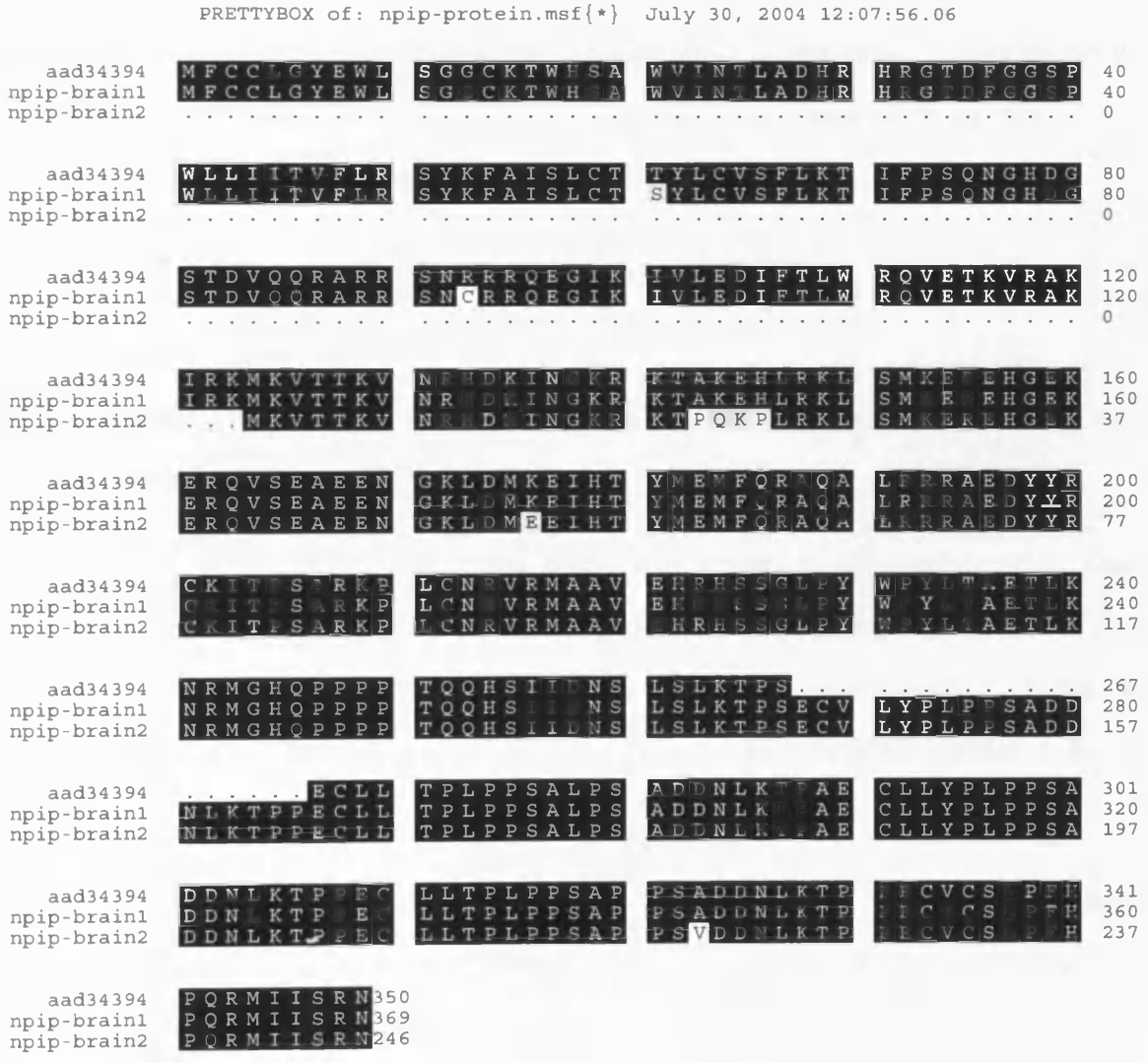


Figure B.7: Multiple alignment of NPIP reference protein AAD34394 with longest open reading frame translations of 2 cDNAs found in brain.

Bibliography

Ahnelt, P. and Kolb, H. (1994). Horizontal cells and cone photoreceptors in human retina: a Golgi-electron microscopic study of spectral connectivity. *J Comp Neurol*, 343(3):406–27.

al Ubaidi, M., Pittler, S., Champagne, M., Triantafyllos, J., McGinnis, J., and Baehr, W. (1990). Mouse opsin. Gene structure and molecular basis of multiple transcripts. *J Biol Chem*, 265(33):20563–9.

Allikmets, R. (1999). Molecular genetics of age-related macular degeneration: current status. *Eur J Ophthalmol*, 9(4):255–65.

Allikmets, R. (2000). Simple and complex ABCR: genetic predisposition to retinal disease. *Am J Hum Genet*, 67(4):793–9.

Allikmets, R., Singh, N., Sun, H., Shroyer, N., Hutchinson, A., Chidambaram, A., Gerrard, B., Baird, L., Stauffer, D., Peiffer, A., Rattner, A., Smallwood, P., Li, Y., Anderson, K., Lewis, R., Nathans, J., Leppert, M., Dean, M., and Lupski, J. (1997). A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat Genet*, 15(3):236–46.

Antos, C., McKinsey, T., Dreitz, M., Hollingsworth, L., Zhang, C., Schreiber, K., Rindt, H., Gorczynski, R., and Olson, E. (2003). Dose-dependent blockade to cardiomyocyte hypertrophy by histone deacetylase inhibitors. *J Biol Chem*, 278(31):28930–7.

Baird, P. N., Guida, E., Chu, D. T., Vu, H. T. V., and Guymer, R. H. (2004). The epsilon2

and epsilon4 alleles of the apolipoprotein gene are associated with age-related macular degeneration. *Invest Ophthalmol Vis Sci*, 45(5):1311–5.

Bauer, W., Hayes, J., White, J., and Wolffe, A. (1994). Nucleosome structural changes due to acetylation. *J Mol Biol*, 236(3):685–90.

Becker-André, M. and Hahlbrock, K. (1989). Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic Acids Res*, 17(22):9437–46.

Bernstein, S., Borst, D., Neuder, M., and Wong (1996). Characterization of a human fovea cDNA library and regional differential gene expression in the human retina. *Genomics*, 32:Genomics 32, 301308.

Bernstein, S., Breding, D., and Fisher, S. (1984). The influence of light on cone disk shedding in the lizard, Sceloporus occidentalis. *J Cell Biol*, 99(2):379–89.

Bernstein, S. and Wong, P. (1998). Regional expression of disease-related genes in human and monkey retina. *Mol Vis*, 4:24.

Bridges, C. (1936). The bar "gene" duplication. *Science 83, 210-211.*

Bustin, S. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol*, 25(2):169–93.

Chapple, J., Hardcastle, A., Grayson, C., Spackman, L., Willison, K., and Cheetham, M. (2000). Mutations in the N-terminus of the X-linked retinitis pigmentosa protein RP2 interfere with the normal targeting of the protein to the plasma membrane. *Hum Mol Genet*, 9(13):1919–26.

Chen, Y., Dougherty, E., and Bittner, M. (1997). Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics*, 2:364–374.

Chevaillier, P. (1993). Pest sequences in nuclear proteins. *Int J Biochem*, 25(4):479–82.

Chong, N., Alexander, R., Gin, T., Bird, A., and Luthert, P. (2000). TIMP-3, collagen, and elastin immunohistochemistry and histopathology of Sorsby's fundus dystrophy. *Invest Ophthalmol Vis Sci*, 41(3):898–902.

Darwin, C. R. (1859). *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* John Murray, London.

David, D., Cardoso, J., Marques, B., Marques, R., Silva, E., Santos, H., and MG, M. B. (2003). Molecular characterization of a familial translocation implicates disruption of hdac9 and possible position effect on tgfbeta2 in the pathogenesis of peters' anomaly. *Genomics*, 81(5):489–503.

den Hollander, A., ten Brink, J., de Kok, Y., van Soest, S., van den Born, L., van Driel, M., van de Pol, D., Payne, A., Bhattacharya, S., Kellner, U., Hoyng, C., Westerveld, A., Brunner, H., Bleeker-Wagemakers, E., Deutman, A., Heckenlively, J., Cremers, F., and Bergen, A. (1999a). Mutations in a human homologue of Drosophila crumbs cause retinitis pigmentosa (RP12). *Nat Genet*, 23(2):217–21.

den Hollander, A., van Driel, M., de Kok, Y., van de Pol, D., Hoyng, C., Brunner, H., Deutman, A., and Cremers, F. (1999b). Isolation and mapping of novel candidate genes for retinal disorders using suppression subtractive hybridization. *Genomics*, 58(3):240–9.

Deutman, A. (1969). Electro-oculography in families with vitelliform dystrophy of the fovea. Detection of the carrier state. *Arch Ophthalmol*, 81(3):305–16.

Diatchenko, L., Lau, Y., Campbell, A., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E., and Siebert, P. (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A*, 93(12):6025–30.

Dobzhansky, T. (1956). *The Biological Basis of Human Freedom.* New York: Columbia Univ. Press.

Dorey, C., Wu, G., Ebenstein, D., Garsd, A., and Weiter, J. (1989). Cell loss in the aging retina. Relationship to lipofuscin accumulation and macular degeneration. *Invest Ophthalmol Vis Sci*, 30(8):1691–9.

Downes, S., Fitzke, F., Holder, G., Payne, A., Bessant, D., Bhattacharya, S., and Bird, A. (1999). Clinical features of codon 172 RDS macular dystrophy: similar phenotype in 12 families. *Arch Ophthalmol*, 117(10):1373–83.

Drasdo, N. and Fowler, C. (1974). Non-linear projection of the retinal image in a wide-angle schematic eye. *Br J Ophthalmol*, 58(8):709–14.

Duggan, D., Bittner, M., Chen, Y., Meltzer, P., and Trent, J. (1999). Expression profiling using cDNA microarrays. *Nat Genet*, 21(1 Suppl):10–4.

Dulai, K., von Dornum, M., Mollon, J., and Hunt, D. (1999). The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates. *Genome Res*, 9(7):629–38.

Dykhuizen, D. and Hartl, D. (1980). Selective neutrality of 6PGD allozymes in E. coli and the effects of genetic background. *Genetics*, 96(4):801–17.

Edwards, A. O., Iii, R. R., Abel, K. J., Manning, A., Panhuysen, C., and Farrer, L. A. (2005). Complement Factor H Polymorphism and Age-Related Macular Degeneration. *Science*.

Eksandh, L., Ponjavic, V., Ayyagari, R., Bingham, E., Hiriyanna, K., Andréasson, S., Ehinger, B., and Sieving, P. (2000). Phenotypic expression of juvenile X-linked retinoschisis in Swedish families with different mutations in the XLRS1 gene. *Arch Ophthalmol*, 118(8):1098–104.

Elledge, S., Mulligan, J., Ramer, S., Spottswood, M., and Davis, R. (1991). Lambda YES: a multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and Escherichia coli mutations. *Proc Natl Acad Sci U S A*, 88(5):1731–5.

Endris, V., Wogatzky, B., Leimer, U., Bartsch, D., Zatyka, M Latif, F., Maher, E., Tariverdian, G., Kirsch, S., Karch, D., and Rappold, G. (2002). The novel rho-gtpase activating gene megap/srgap3 has a putative role in severe mental retardation. *Proc Nat Acad Sci*, 99:11754–11759.

Evans, W. and Wormald, R. (1996). Is the incidence of registrable age-related macular degeneration increasing? *Br J Ophthalmol*, 80:9–14.

Fain, G., Matthews, H., Cornwall, M., and Koutalos, Y. (2001). Adaptation in vertebrate photoreceptors. *Physiol Rev*, 81:Physiol Rev 81, 117–151.

Ferreira, P., Nakayama, T., Pak, W., and Travis, G. (1996). Cyclophilin-related protein RanBP2 acts as chaperone for red/green opsin. *Nature*, 383(6601):637–40.

Finckh, U., Xu, S., Kumaramanickavel, G., Schürmann, M., Mukkadan, J., Fernandez, S., John, S., Weber, J., Denton, M., and Gal, A. (1998). Homozygosity mapping of autosomal recessive retinitis pigmentosa locus (RP22) on chromosome 16p12.1-p12.3. *Genomics*, 48(3):341–5.

Finkelstein, D., Ewing, R., Gollub, J., Sterky, F., Cherry, J. M., and Somerville, S. (2002). Microarray data quality analysis: lessons from the AFGC project. Arabidopsis Functional Genomics Consortium. *Plant Mol Biol*, 48(1-2):119–31.

Fish, G., Grey, R., Sehmi, K., and Bird, A. (1981). The dark choroid in posterior retinal dystrophies. *Br J Ophthalmol*, 65(5):359–63.

Fodor, S., Read, J., Pirrung, M., Stryer, L., Lu, A., and Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science*, 251(4995):767–73.

Fuchs, S., Nakazawa, M., Maw, M., Tamai, M., Oguchi, Y., and Gal, A. (1995). A homozygous 1-base pair deletion in the arrestin gene is a frequent cause of oguchi disease in japanese. *Nature Genet*, 10:360–362.

George, N., Yates, J., Bradshaw, K., and Moore, A. (1995a). Infantile presentation of X linked retinoschisis. *Br J Ophthalmol*, 79(7):653–7.

George, N., Yates, J., and Moore, A. (1995b). X linked retinoschisis. *Br J Ophthalmol*, 79(7):697–702.

Gillespie, D. and Spiegelman, S. (1965). A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane. *J Mol Biol*, 12(3):829–42.

Ginzinger, D. G. (2002). Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol*, 30(6):503–12.

Goodman, M., Moore, G., and Matsuda, G. (1975). Darwinian evolution in the genealogy of haemoglobin. *Nature*, 253(5493):603–8.

Gorin, M., Breitner, J., Jong, P. D., Hageman, G., Klaver, C., Kuehn, M., and Seddon, J. (1999). The genetics of age-related macular degeneration. *Mol Vis*, 5:29.

Gottfredsdottir, M., Sverrisson, T., Musch, D., and Stefánsson, E. (1999). Age related macular degeneration in monozygotic twins and their spouses in Iceland. *Acta Ophthalmol Scand*, 77(4):422–5.

Grauer, D. and Li, W.-H. (2000). *Fundamentals of Molecular Evolution, 2nd Edition.* Sinauer Associates.

Group, A.-R. E. D. S. R. (2001). A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8. *Arch Ophthalmol*, 119(10):1417–36.

Haines, J. L., Hauser, M. A., Schmidt, S., Scott, W. K., Olson, L. M., Gallins, P., Spencer, K. L., Kwan, S. Y., Noureddine, M., Gilbert, J. R., Schnetz-Boutaud, N., Agarwal, A., Postel, E. A., and Pericak-Vance, M. A. (2005). Complement Factor H Variant Increases the Risk of Age-Related Macular Degeneration. *Science*.

Haldane, J. (1932). *The causes of evolution.*

Haley, T., Pochet, R., Baizer, L., Burton, M., Crabb, J., Parmentier, M., and Polans, A. (1995). Calbindin D-28K immunoreactivity of human cone cells varies with retinal position. *Vis Neurosci*, 12(2):301–7.

Harding, H., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Shapira, M., and Ron, D. (2000). Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Molecular Cell*, 6:1099–1108.

Heiba, I., Elston, R., Klein, B., and Klein, R. (1994). Sibling correlations and segregation analysis of age-related maculopathy: the beaver dam eye study. *Genet Epidemiol*, 11:51–67.

Hendrickson, A. (1994). Primate foveal development: A microcosm of current questions in neurobiology. *Invest Ophthalmol Visual Sci*, 35:3129 3133.

Higuchi, R., Fockler, C., Dollinger, G., and Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y)*, 11(9):1026–30.

Huber, W., von Heydebreck, A., Sultmann, H., Poutska, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18:S1 S96–S104.

Hughes, A. (1994). The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B Biol Sci*, 256(1346):119–24.

Hurwitz, R., Bunt-Milam, A., and Beavo, J. (1984a). Immunologic characterization of the photoreceptor outer segment cyclic GMP phosphodiesterase. *J Biol Chem*, 259(13):8612–8.

Hurwitz, R., Hansen, R., Harrison, S., Martins, T., Mumby, M., and Beavo, J. (1984b).

Immunologic approaches to the study of cyclic nucleotide phosphodiesterases. *Adv Cyclic Nucleotide Protein Phosphorylation Res*, 16:89–106.

Ibbotson, R., Hunt, D., Bowmaker, J., and Mollon, J. (1992). Sequence divergence and copy number of the middle- and long-wave photopigment genes in Old World monkeys. *Proc R Soc Lond B Biol Sci*, 247(1319):145–54.

Iyengar, S. K., Song, D., Klein, B. E. K., Klein, R., Schick, J. H., Humphrey, J., Millard, C., Liptak, R., Russo, K., Jun, G., Lee, K. E., Fijal, B., and Elston, R. C. (2004). Dissection of genomewide-scan data in extended families reveals a major locus and oligogenic susceptibility for age-related macular degeneration. *Am J Hum Genet*, 74(1):20–39.

Jacobson, S., Cideciyan, A., Regunath, G., Rodriguez, F., Vandenburgh, K., Sheffield, V., and Stone, E. (1995). Night blindness in Sorsby's fundus dystrophy reversed by vitamin A. *Nat Genet*, 11(1):27–32.

Johnson, M., Viggiano, L., Bailey, J., Abdul-Rauf, M., Goodwin, G., Rocchi, M., and Eichler, E. (2001). Positive selection of a gene family during the emergence of humans and African apes. *Nature*, 413(6855):514–9.

Kafatos, F., Jones, C., and Efstratiadis, A. (1979). Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res*, 7(6):1541–52.

Kajiwara, K., Hahn, L., Mukai, S., Travis, G., Berson, E., and Dryja, T. (1991). Mutations in the human retinal degeneration slow gene in autosomal dominant retinitis pigmentosa. *Nature*, 354(6353):480–3.

Kemp, C., Jacobson, S., Cideciyan, A., Kimura, A., Sheffield, V., and Stone, E. (1994). RDS gene mutations causing retinitis pigmentosa or macular degeneration lead to the same abnormality in photoreceptor function. *Invest Ophthalmol Vis Sci*, 35(8):3154–62.

Kennan, A., Aherne, A., Palfi, A., Humphries, M., McKee, A., Stitt, A., Simpson, D., Demtroder, K., Orntoft, T., Ayuso, C., Kenna, P., Farrar, G., and Humphries, P. (2002). Identification of an impdh1 mutation in autosomal dominant retinitis pigmentosa (rp10) revealed following comparative microarray analysis of transcripts derived from retinas of wild-type and rho(-/-) mice. *Hum Mol Genet*, 11(5):547–57.

Khrapko, K., YuP, K. L., Khorlyn, A., Shick, V., Florentiev, V., and Mirzabekov, A. (1989). An oligonucleotide hybridization approach to DNA sequencing. *FEBS Lett*, 256(1-2):118–22.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(129):624–6.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge Univ. Press.

Klaver, C., Wolfs, R., Assink, J., van Duijn, C., Hofman, A., and de Jong, P. (1998). Genetic risk of age-related maculopathy population-based familial aggregation study. *Arch Ophthalmol*, 116:1646–1651.

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J.-Y., Sackler, R. S., Haynes, C., Henning, A. K., Sangiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C., and Hoh, J. (2005). Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science*.

Knight, J. (2001). When the chips are down. *Nature*, 410(6831):860–1.

Kolb, H., Fernandez, E., Schouten, J., Ahnelt, P., Linberg, K., and Fisher, S. (1994). Are there three types of horizontal cell in the human retina? *J Comp Neurol*, 343(3):370–86.

Kolb, H., Linberg, K., and Fisher, S. (1992). Neurons of the human retina: a golgi study. *J Comp Neurol*, 31:147–187.

Kopito, R. R. (2003). The missing linker: an unexpected role for a histone deacetylase. *Mol Cell*, 12(6):1349–51.

Kremer, H., Pinckers, A., van den Helm, B., Deutman, A., Ropers, H., and Mariman, E. (1994). Localization of the gene for dominant cystoid macular dystrophy on chromosome 7p. *Hum Mol Genet*, 3(2):299–302.

Lee, D., Hayes, J., Pruss, D., and Wolffe, A. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell*, 72(1):73–84.

Lee, W. (1997). *Molecular Evolution*. Sinauer, Sunderland.

Lennon, G. and Lehrach, H. (1991). Hybridization analyses of arrayed cDNA libraries. *Trends Genet*, 7(10):314–7.

Livak, K. and Schmittgen, T. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25(4):402–8.

Livesey, F., Furukawa, T., MA, S., GM, C., and CL, C. (2000). Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene crx. *Curr Biol*, 10(6):301–10.

Majid, M. A., Smith, V. A., Easty, D. L., Baker, A. H., and Newby, A. C. (2002). Adenovirus mediated gene delivery of tissue inhibitor of metalloproteinases-3 induces death in retinal pigment epithelial cells. *Br J Ophthalmol*, 86(1):97–101.

Manschot, W. (1972). Pathology of hereditary juvenile retinoschisis. *Arch Ophthalmol*, 88(2):131–8.

Mariani, A. (1990). Amacrine cells of the rhesus monkey retina. *J Comp Neurol*, 301(3):382–400.

Marmorstein, A., Marmorstein, L., Rayborn, M., Wang, X., Hollyfield, J., and Petrukhin, K. (2000). Bestrophin, the product of the Best vitelliform macular dystrophy gene (VMD2), localizes to the basolateral plasma membrane of the retinal pigment epithelium. *Proc Natl Acad Sci U S A*, 97(23):12758–63.

Marmorstein, L. Y., Munier, F. L., Arsenijevic, Y., Schorderet, D. F., McLaughlin, P. J., Chung, D., Traboulsi, E., and Marmorstein, A. D. (2002). Aberrant accumulation of EFEMP1 underlies drusen formation in Malattia Leventinese and age-related macular degeneration. *Proc Natl Acad Sci U S A*, 99(20):13067–72.

Martínez-Mir, A., Paloma, E., Allikmets, R., Ayuso, C., del Rio, T., Dean, M., Vilageliu, L., Gonzàlez-Duarte, R., and Balcells, S. (1998). Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. *Nat Genet*, 18(1):11–2.

Maugeri, A., Klevering, B., Rohrschneider, K., Blankenagel, A., Brunner, H., Deutman, A., Hoyng, C., and Cremers, F. (2000). Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal recessive cone-rod dystrophy. *Am J Hum Genet*, 67(4):960–6.

McKechnie, N., Al-Mahdawi, S., Dutton, G., and Forrester, J. (1986). Ultrastructural localization of retinal s antigen in the human retina. *Exp Eye Res*, 42(5):479–87.

Meyers, S., Greene, T., and Gutman, F. (1995). A twin study of age-related macular degeneration. *Am J Ophthalmol*, 120(6):757–66.

Mirshahi, M., Thillaye, B., Tarraf, M., de Kozak, Y., and Faure, J. (1994). Light-induced changes in s-antigen (arrestin) localization in retinal photoreceptors: differences between rods and cones and defective process in rcs rat retinal dystrophy. *Eur J Cell Biol*, 63(1):61–7.

Mohler, C. and Fine, S. (1981). Long-term evaluation of patients with Best's vitelliform dystrophy. *Ophthalmology*, 88(7):688–92.

Molday, L., Hicks, D., Sauer, C., Weber, B., and Molday, R. (2001). Expression of X-linked retinoschisis protein RS1 in photoreceptor and bipolar cells. *Invest Ophthalmol Vis Sci*, 42(3):816–25.

Molday, L., Rabin, A., and Molday, R. (2000). ABCR expression in foveal cone photoreceptors and its role in Stargardt macular dystrophy. *Nat Genet*, 25(3):257–8.

Müller, H., Prokofyeva, A., and Raffel, D. (1935). Minute intergenic rearrangement as a cause of apparent "gene mutation". *Nature 135:253-255.*

Nadon, R. and Shoemaker, J. (2002). Statistical issues with microarrays: processing and analysis. *Trends Genet*, 18(5):265–71.

Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., and Ohara, O. (1998). Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res*, 5(5):277–86.

Nichols, B., Sheffield, V., Vandenburgh, K., Drack, A., Kimura, A., and Stone, E. (1993). Butterfly-shaped pigment dystrophy of the fovea caused by a point mutation in codon 167 of the RDS gene. *Nat Genet*, 3(3):202–7.

Nork, T., Mangini, N., and Millecchia, L. (1993). Rods and cones contain antigenically distinctive s-antigens. *Invest Ophthalmol Vis Sci*, 34(10):2918–25.

Ohno, S. (1970). *Evolution by Gene Duplication.* Springer, New York.

O'Shea, J. (1998). Age-related macular degeneration. *Postgrad Med J*, 74(870):203–7.

Peirson, S., Butler, J., and Foster, R. (2003). Experimental validation of novel and conventional approaches to quantitative real-time pcr data analysis. *Nucleic Acids Research*, 31(14):e73.

Petrie, K., Guidez, F., Howell, L., Healy, L., Waxman, S., Greaves, M., and Zelent, A. (2003). The histone deacetylase 9 gene encodes multiple protein isoforms. *J Biol Chem*, 2(18):16059–72.

Petrukhin, K., Koisti, M., Bakall, B., Li, W., Xie, G., Marknell, T., Sandgren, O., Forsman, K., Holmgren, G., Andreasson, S., Vujic, M., Bergen, A., McGarty-Dugan, V., Figueroa, D., Austin, C., Metzker, M., Caskey, C., and Wadelius, C. (1998). Identification of the gene responsible for Best macular dystrophy. *Nat Genet*, 19(3):241–7.

Pfaffl, M. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*, 29(9):e45.

Pianta, M. J., Aleman, T. S., Cideciyan, A. V., Sunness, J. S., Li, Y., Campochiaro, B. A., Campochiaro, P. A., Zack, D. J., Stone, E. M., and Jacobson, S. G. (2003). In vivo micropathology of Best macular dystrophy with optical coherence tomography. *Exp Eye Res*, 76(2):203–11.

Piatigorsky, J., O'Brien, W., Norman, B., Kalumuck, K., Wistow, G., Borras, T., Nickerson, J., and Wawrousek, E. (1988). Gene sharing by delta-crystallin and argininosuccinate lyase. *Proc Natl Acad Sci U S A*, 85(10):3479–83.

Poetsch, A., Molday, L., and Molday, R. (2001). The cGMP-gated channel and related glutamic acid-rich proteins interact with peripherin-2 at the rim region of rod photoreceptor disc membranes. *J Biol Chem*, 276(51):48009–16.

Polyak, S. (1941). *The Retina*. University of Chicago Press, Chicago.

Qi, J. H., Ebrahem, Q., Moore, N., Murphy, G., Claesson-Welsh, L., Bond, M., Baker, A., and Anand-Apte, B. (2003). A novel function for tissue inhibitor of metalloproteinases-3 (TIMP3): inhibition of angiogenesis by blockage of VEGF binding to VEGF receptor-2. *Nat Med*, 9(4):407–15.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nat Genet*, 32 Suppl:496–501.

Reid, D., KU, L., Campbell, A., and Forrester, J. (1987). Electron immunocytochemical localization of retinal s-antigen with a rat monoclonal antibody. *Exp Eye Res*, 45(5):731–45.

Rodieck, R. (1998). *The First Steps in Seeing*. Sinauer Associates, Inc. Sunderland, Massachusetts.

Rogers, S., Wells, R., and Rechsteiner, M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*, 234(4774):364–8.

Ross, D., Scherf, U., Eisen, M., Perou, C., Rees, C., Spellman, P., Iyer, V., Jeffrey, S., de Rijn, M. V., Waltham, M., Pergamenschikov, A., Lee, J., Lashkari, D., Shalon, D., Myers, T., Weinstein, J., Botstein, D., and Brown, P. (2000). Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet*, 24(3):227–35.

Salminen, A., Tapiola, T., Korhonen, P., and Suuronen, T. (1998). Neuronal apoptosis induced by histone deacetylase inhibitors. *Brain Res Mol Brain Res*, 61(1-2):203–6.

Sambrook, J., Fritsch, E., and Manniatis., T. (1989). *Molecular Cloning: A Laboratory Manual, 2nd edition.* Cold Spring Harbor Laboratory Press.

Sauer, C., Gehrig, A., Warneke-Wittstock, R., Marquardt, A., Ewing, C., Gibson, A., Lorenz, B., Jurklies, B., and Weber, B. (1997). Positional cloning of the gene associated with X-linked juvenile retinoschisis. *Nat Genet*, 17(2):164–70.

Schena, M., Shalon, D., Davis, R., and Brown, P. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235):467–70.

Schick, J., Iyengar, S., Klein, B., Klein, R., Reading, K., Liptak, R., Millard, C., Lee, K., Tomany, S., Moore, E., Fijal, B., and Elston, R. (2003). A whole-genome screen of a quantitative trait of age-related maculopathy in sibships from the beaver dam eye study. *Am J Hum Genet*, 72(6):1412–24.

Schmidt, S., Scott, W., Postel, E., Agarwal, A., Hauser, E., Paz, M. D. L., Gilbert, J., Weeks, D., Gorin, M., Haines, J., and Pericak-Vance, M. (2004). Ordered subset linkage analysis supports a susceptibility locus for age-related macular degeneration on chromosome 16p12. *BMC Genet*, 5(1):18.

Schultz, D. W., Klein, M. L., Humpert, A. J., Luzier, C. W., Persun, V., Schain, M., Mahan, A., Runckel, C., Cassera, M., Vittal, V., Doyle, T. M., Martin, T. M., Weleber,

R. G., Francis, P. J., and Acott, T. S. (2003). Analysis of the ARMD1 locus: evidence that a mutation in HEMICENTIN-1 is associated with age-related macular degeneration in a large family. *Hum Mol Genet*, 12(24):3315–23.

Sharon, D., Blackshaw, S., Cepko, C., and Dryja, T. (2002). Profile of the genes expressed in the human peripheral retina, macula, and retinal pigment epithelium determined through serial analysis of gene expression (SAGE). *Proc Natl Acad Sci*, 99(1):315–320.

Smyth, G. K. and Speed, T. (2003). Normalization of cDNA microarray data. *Methods*, 31(4):265–73.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*, 98(3):503–17.

Southern, E., Mir, K., and Shchepinov, M. (1999). Molecular interactions on microarrays. *Nat Genet*, 21(1 Suppl):5–9.

Stone, E., Lotery, A., Munier, F., Héon, E., Piguet, B., Guymer, R., Vandenburgh, K., Cousin, P., Nishimura, D., Swiderski, R., Silvestri, G., Mackey, D., Hageman, G., Bird, A., Sheffield, V., and Schorderet, D. (1999). A single EFEMP1 mutation associated with both Malattia Leventinese and Doyne honeycomb retinal dystrophy. *Nat Genet*, 22(2):199–202.

Stone, E. M., Braun, T. A., Russell, S. R., Kuehn, M. H., Lotery, A. J., Moore, P. A., Eastman, C. G., Casavant, T. L., and Sheffield, V. C. (2004). Missense variations in the fibulin 5 gene and age-related macular degeneration. *N Engl J Med*, 351(4):346–53.

Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J. P., Gunter, B., Linseley, P. S., Stoughton, R. B., Aebersold, R., Collins, S. J., Hanlon, W. A., and Hood, L. E. (2004). Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Mol Cell Proteomics*.

Travis, G., Sutcliffe, J., and Bok, D. (1991). The retinal degeneration slow (rds) gene product is a photoreceptor disc membrane-associated glycoprotein. *Neuron*, 6(1):61–70.

Tricarico, C., Pinzani, P., Bianchi, S., Paglierani, M., Distante, V., Pazzagli, M., Bustin, S., and Orlandoa, C. (2002). Quantitative real-time reverse transcription polymerase chain reaction: normalization to rrna or single housekeeping genes is inappropriate for human tissue. *Analytical Biochemistry*, 309:Analytical Biochemistry 309 293300.

Tseng, G., Oh, M., Rohlin, L., Liao, J., and Wong, W. (2001). Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res*, 29(12):2549–57.

Vacquier, V., Swanson, W., and Lee, Y. (1997). Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? *J Mol Evol*, 44 Suppl 1:S15–22.

Vandesompele, J., De Preter, K., Pattyn, P., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative rt-pcr data by geometric averaging of multiple internal control genes. *Genome Biology*, 3:7 research0034100341.

Velculescu, V., Zhang, L., Vogelstein, B., and Kinzler, K. (1995). Serial analysis of gene expression. *Science*, 270(5235):484–7.

Walker, N. J. (2002). Tech.Sight. A technique whose time has come. *Science*, 296(5567):557–9.

Warrington, J., Nair, A., Mahadevappa, M., and Tsyganskaya, M. (2000). Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics*, 2(3):143–7.

Weber, B., Vogt, G., Pruett, R., Stöhr, H., and Felbor, U. (1994). Mutations in the tissue

inhibitor of metalloproteinases-3 (TIMP3) in patients with Sorsby's fundus dystrophy. *Nat Genet*, 8(4):352–6.

Weber, B. H. F., Schrewe, H., Molday, L. L., Gehrig, A., White, K. L., Seeliger, M. W., Jaissle, G. B., Friedburg, C., Tamm, E., and Molday, R. S. (2002). Inactivation of the murine X-linked juvenile retinoschisis gene, Rs1h, suggests a role of retinoschisin in retinal cell layer organization and synaptic structure. *Proc Natl Acad Sci U S A*, 99(9):6222–7.

Weeks, D., Conley, Y., Mah, T., Paul, T., Morse, L., Ngo-Chang, J., Dailey, J., Ferrell, R., and Gorin, M. (2000). A full genome scan for age-related maculopathy. *Hum Mol Genet*, 22(9):1329–49.

Weeks, D., Conley, Y., Tsai, H., Mah, T., Rosenfeld, P., Paul, T., Eller, A., Morse, L., Dailey, J., Ferrell, R., and Gorin, M. (2001). Age-related maculopathy: an expanded genome-wide scan with evidence of susceptibility loci within the 1q31 and 17q25 regions. *Am J Ophthalmol*, 132(5):682–92.

Weeks, D. E., Conley, Y. P., Tsai, H.-J., Mah, T. S., Schmidt, S., Postel, E. A., Agarwal, A., Haines, J. L., Pericak-Vance, M. A., Rosenfeld, P. J., Paul, T. O., Eller, A. W., Morse, L. S., Dailey, J. P., Ferrell, R. E., and Gorin, M. B. (2004). Age-Related Maculopathy: A Genomewide Scan with Continued Evidence of Susceptibility Loci within the 1q31, 10q26, and 17q25 Regions. *Am J Hum Genet*, 75(2).

Weingeist, T., Kobrin, J., and Watzke, R. (1982). Histopathology of Best's macular dystrophy. *Arch Ophthalmol*, 100(7):1108–14.

Weng, J., Mata, N., Azarian, S., Tzekov, R., Birch, D., and Travis, G. (1999). Insights into the function of Rim protein in photoreceptors and etiology of Stargardt's disease from the phenotype in abcr knockout mice. *Cell*, 98(1):13–23.

Winderickx, J., Battisti, L., Hibiya, Y., Motulsky, A., and Deeb, S. (1993). Haplotype

diversity in the human red and green opsin genes: evidence for frequent sequence exchange in exon 3. *Hum Mol Genet*, 2(9):1413–21.

Yang, I. V., Chen, E., Hasseman, J. P., Liang, W., Frank, B. C., Wang, S., Sharov, V., Saeed, A. I., White, J., Li, J., Lee, N. H., Yeatman, T. J., and Quackenbush, J. (2002a). Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome Biol*, 3(11):research0062.

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002b). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*, 30(4):e15.

Yates, J. and Moore, A. (2000). Genetic susceptibility to age related macular degeneration. *J Med Genet*, 37(2):83–7.

Young, R. (1967). The renewal of photoreceptor cell outer segments. *J Cell Biol*, 33:61–72.

Zhang, C., McKinsey, T., Chang, S., Antos, C., Hill, J., and Olson, E. (2002). Class ii histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell*, 110(4):479–88.

Zhou, X., Marks, P., Rifkind, R., and Richon, V. (2001). Cloning and characterization of a histone deacetylase, hdac9. *Proc Nat Acad Sci*, 98:10572–10577.

Zuckerland, E. and Pauling, L. (1965). *Evolving Genes and Proteins*. Academic, New York. pp. 97-166.

Expression of *PRPF31* mRNA in Patients with Autosomal Dominant Retinitis Pigmentosa: A Molecular Clue for Incomplete Penetrance?

Eranga N. Vithana,[1,2] *Leen Abu-Safieh,*[1,2] *Lucia Pelosini,*[3] *Elizabeth Winchester,*[3] *Dan Hornan,*[1] *Alan C. Bird,*[3] *David M. Hunt,*[1] *Stephen A. Bustin,*[4] *and Shomi S. Bhattacharya*[1]

Investigative Ophthalmology & Visual Science, October 2003, Vol. 44, No. 10

IOVS, October 2003, Vol. 44, No. 10